

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

The Nebraska Educator: A Student-Led Journal Department of Teaching, Learning and Teacher Education

10-2020

The Nebraska Educator, Volume 5: 2020

College of Education and Human Sciences University of Nebraska-Lincoln

Follow this and additional works at: <https://digitalcommons.unl.edu/nebeducator>



Part of the [Teacher Education and Professional Development Commons](#)

College of Education and Human Sciences University of Nebraska-Lincoln, "The Nebraska Educator, Volume 5: 2020" (2020). *The Nebraska Educator: A Student-Led Journal*. 44.
<https://digitalcommons.unl.edu/nebeducator/44>

This Article is brought to you for free and open access by the Department of Teaching, Learning and Teacher Education at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in The Nebraska Educator: A Student-Led Journal by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

The Nebraska Educator

Volume 5: 2020

Published by the College of Education and Human Sciences
University of Nebraska-Lincoln

Editor-in-Chief

Taylor Hamblin

Senior Editors

Justin Andersson

Amy Zachek

Associate Editors

Sarah Garvey

Olivia Kennedy

ISSN 2375-6853

doi: 10.32873/unl.dc.ne000

The Nebraska Educator

nebeducator@unl.edu

College of Education and Human Sciences

University of Nebraska-Lincoln

Lincoln, Nebraska 68588-0355

<https://cehs.unl.edu/cehs/student-organizations/nebraska-educator/>

<http://digitalcommons.unl.edu/nebeducator/>



© 2020 University of Nebraska-Lincoln

Contents

- 4 EmPOWER: An Adaptable Writing Intervention
 Carly Dinnes, Teaching, Learning, and Teacher Education
- 25 Learning to Lesson Plan: A Mentor’s Impact on Pre-service Teachers
 Kelly M. Gomez, Teaching, Learning, and Teacher Education
- 57 Difficulty and Distance in Educational Encounters with Historical Violence
 Grant Scribner, Teaching, Learning, and Teacher Education
- 70 Culturally Relevant Science Teaching: A Literature Review
 Uma Ganesan, Teaching, Learning, and Teacher Education
- 92 Interval Estimation of Proportion of Second-level Variance in Multi-level Modeling
 Steven Svoboda, Education Psychology
- 105 The History, Evolution, and Trends of Academic Dishonesty: A Literature Review
 Amy Zachek, Educational Administration
- 121 Bilingual Acculturation Assessment: An Overview of Current Developments
 Donna Chen, Educational Psychology
- 148 The Current State of Assessing Historical Thinking: A Literature Analysis
 Taylor Hamblin, Teaching, Learning, and Teacher Education
- 177 Language and Power
 Alessia Barbici-Wagner, Teaching, Learning, and Teacher Education
- 194 Seeing Formative Assessments from a Broad Perspective
 Consuelo M. Gallardo, Teaching, Learning, and Teacher Education
- 216 Resources for Content-area Teachers Educating ELLs: A Literature Review
 Heidi Bartlett, Teaching, Learning, and Teacher Education

Letter From The Editor

When writing an editor's note, one naturally thinks about the past and tries to unpack meaning from previous challenges and successes, and hopefully in the process of reminiscing, illuminates the toils done by those who made the achievement possible. This year, the accomplishment is a volume of *The Nebraska Educator* which houses eleven publications, a record in this journal's history. Along with an impressive quantity, volume V contains exceptional, diverse articles from multiple departments in the College of Education and Human Sciences (CEHS) at the University of Nebraska-Lincoln. To deduce the reason behind this success, I recollected this last year, and what consistently came to mind was the editorial board members who inspired me due to their pursuit of exceptionalism.

Between the Fall 2019 semester until now, a time many will remember as being volatile, a storm full of tragedy, working with my fellow editors has been my silver lining in the gray clouds that billowed forth. The size and quality of the most recent volume could not have occurred if it were not for the outstandingly dedicated graduate students who possessed a true spirit of leadership and volunteerism. I have worked in student organizations for over twelve years, and I have never witnessed a group of volunteer leaders so committed to going beyond the average in their pursuit of the extraordinary. Due to the efforts of the editorial board, *The Nebraska Educator* is producing a volume for the first time in three years, possesses a constitution, increased editorial board membership that represents a multitude of departments within CEHS, and even now, these graduate students continue to find ways to provide more opportunities for their peers to publish research.

Along with the commitment of editorial board members, this publication would not be possible without the guidance of CEHS faculty and staff, graduate students who served as external reviewers, and Sue Gardner and Paul Royster, who helped publish the journal on Digital Commons. I especially would like to thank Guy Trainin, who was an early source of insight and a tolerant listener of novel ideas. The financial assistance from the department of Teaching, Learning, and Teacher Education was incredibly helpful in promoting the journal's purpose.

It has been our utmost pleasure in providing a place for aspiring research scholars to publish their work. The articles *The Nebraska Educator* publishes have been downloaded tens of thousands of times; it is our sincerest belief that the eleven new additions will continue to push many academic conversations forward. I hope you enjoy the following publications as much as I took pleasure in working with the remarkable people that brought them together.

Respectfully yours,
Taylor Hamblin
Editor-in-Chief, 2020

The editors of *The Nebraska Educator* would like to express our sincere gratitude to the following individuals who lent their time and talents to serve as external reviewers for this volume. We would not be possible without your tireless efforts.

Fitsum Adebe
James E. Baker
Trenton Haltom
Elizabeth Hasseler
Uma Ganesan
Scott Gealy
Sophia Knudson
Chris Labenz
Julia Larsen
Xianquan Liu
Junrong Lu
Kelsey Quigley
Hamza Rfssa
Cassandra Schroeder
Grant (Roger) Scribner
Kali Sloup
Amy Sokoll
Elizabeth Svoboda

EmPOWER: An Adaptable Writing Intervention

Carly Dinnes

Department of Teaching, Learning, and Teacher Education
University of Nebraska-Lincoln

Abstract

EmPOWER is a six-stage writing intervention designed by speech-language pathologists to improve the expository writings of school-aged children with language learning and executive function disabilities. The intervention uses scaffolded instruction to transform struggling students into independent and self-regulating writers by training the students to use a variety of supports (e.g., graphic organizers, checklists) and strategies (e.g., referring back to the writing prompt) throughout the writing process. Many key features of the EmPOWER approach to writing instruction directly support components described in cognitive models of writing, which indicates that EmPOWER is a theory-guided writing intervention that may benefit a wide range of individuals beyond just school-aged children. Thus, this analysis provides an overview of the EmPOWER approach, compares the intervention's key features to the most recent conceptualization of the cognitive mechanisms involved in the writing process, and provides a discussion for adapting EmPOWER to a specific population (i.e., people with aphasia).

Keywords: writing, intervention, instruction, theory, aphasia

doi: 10.32873/unl.dc.ne001

Writing is a complex form of communication that requires the recruitment and coordination of multiple skills to effectively convey a meaningful and concise message. Multiple cognitive skills (e.g., recalling relevant information, organizing the information into a meaningful sequence, and evaluating the quality of the message generated) must be employed in tandem with multiple linguistic skills (e.g., retrieving specific words from an internal lexicon to form a message, applying syntactic knowledge to guide message composition, and reading the message in order to identify errors and ambiguities) throughout the writing process to ensure effective communication between writer and reader. The inherent complexity of this method of communication makes writing a difficult skill to master and susceptible to the effects of cognitive and linguistic impairments resulting from developmental or acquired disorders.

Researchers in the fields of education, special education, and speech-language pathology have explored various ways of providing effective writing instruction and intervention for students with and without impairments. Several researchers specifically in the fields of special education and education have developed writing models that detail just how the writing process works with the goal of informing instructional and intervention practices (for examples of writing models, see Graham, 2018; Hayes & Flower, 1980; Kellogg, 1996; Scardamalia & Bereiter, 1987; and Zimmerman & Risemberg, 1997). In addition to theoretical models of the writing process, the writing literature contains multiple experimental research studies that have led to the development of a variety of recommendations for scaffolding writing development such as providing students with models of writing to review, more opportunities to write, and the option to write either by hand or using a computer (Graham & Perin, 2007a, 2007b; Graham, Harris, & Chambers, 2016; Graham, Harris, & Santangelo, 2015; Mason & Graham, 2008; National Commission on Writing, 2003; National Commission on Writing, 2006; Rogers &

Graham, 2008). Also included in the literature is a wide variety of specific educational and intervention strategies that detail more structured approaches to developing and supporting students' written expression (e.g., Self-Regulated Strategy Development, EmPOWER; Harris, Graham, & Mason, 2003; Bashir & Singer, 2006; Singer & Bashir, 2004).

Significantly less literature is available on adult writing. This is a problem for professionals working with adults who have acquired writing difficulties as a result of injuries such as cerebrovascular accident (CVA; i.e., stroke) or traumatic brain injury (TBI). Recent studies identified writing challenges after TBI as having a negative impact on engagement in writing for academic (e.g., report writing, note taking), vocational (e.g., work documentation), social (e.g., email, text messaging), and personal (e.g., journaling, reminder notes) purposes (Dinnes & Hux, 2018; Dinnes, Hux, Holmen, Martens, & Smith, 2018; Ledbetter, Sohlberg, Fickas, Horney, & McIntosh, 2017). This pervasive interference is more pronounced for people with aphasia.

Aphasia is a language disorder frequently caused by CVA and is the source of persistent difficulty with expressive language—such as speaking or writing—as well as comprehending written and spoken information. Much of the current literature on adult writing interventions focuses on people with aphasia; however, these studies investigate the effectiveness of interventions to improve the generation of single words or simple sentences for people struggling with severe writing impairments. Thus, scant research is currently available to provide intervention recommendations for people with less severe acquired writing difficulties.

The scarcity of writing intervention literature for adults struggling with effective written communication necessitates the reference and adaptation of writing instruction and intervention recommendations developed for use with school-age children. Ideally, writing models should

guide both the development and adaptation of writing interventions in order to maximize the efficacy and potency of the intervention by ensuring that multiple, if not all, aspects of the process are directly targeted by the intervention; however, not all researchers and clinicians consider referencing models when developing interventions. Thus, the purpose of this analysis is to (a) examine EmPOWER, a structured approach to writing intervention, (b) compare EmPOWER to a new writing model to identify areas of the writing process that are and are not directly targeted by EmPOWER, and (c) provide a discussion on how EmPOWER can be adapted through modifications and supplementations to become an effective writing intervention for people with aphasia.

EmPOWER

EmPOWER is an intervention that treats writing as a problem-solving task and, thus, specifically targets the various language and executive function skills that underlie the writing process (Bashir & Singer, 2006; Singer & Bashir, 2004). This strategy is an adaptation and expansion of the POWER strategy (Englert et al., 1988), which prompted students with learning disabilities to Plan, Organize, Write, Edit, and Revise. EmPOWER, however, was specifically designed to support students with language and executive function difficulties (Bashir & Singer, 2006; Singer & Bashir, 2004). As such, EmPOWER is a self-regulation approach that teaches writers to employ strategies to assist in the identification and achievement of writing goals using six steps: Evaluate, make a Plan, Organize, Work, Evaluate, and Re-work (Bashir & Singer, 2006; Singer & Bashir, 2004). This approach is implemented as a form of scaffolded instruction in which less support is provided as students begin mastering the various components of the strategy (Bashir & Singer, 2006; Singer & Bashir, 2004). Given sufficient time and practice, students will be able to independently use the EmPOWER approach to support their writing

(Bashir & Singer, 2006; Singer & Bashir, 2004). Although EmPOWER was originally designed by speech-language pathologists (SLPs) to address the expository writing needs of school-aged children with executive function disorders and language learning disabilities, the approach can be adapted for use across a variety of conditions and with a variety of individuals (Bashir & Singer, 2006; Singer & Bashir, 2004). Furthermore, teachers can easily adopt and adapt EmPOWER to suit their needs in the classroom.

The Evaluate stage of EmPOWER cues students to verify their understanding of the writing prompt by identifying specific instructions and segmenting the prompt into separate components (Bashir & Singer, 2006; Singer & Bashir, 2004). Students are taught to identify the action words embedded in the prompt (e.g., describe, tell, explain) and then use this information to determine the number of components their written response must have. For example, students given the prompt, “Select an important figure from this unit, describe their achievements, and explain why their work was important,” would identify three action words (i.e., “select”, “describe”, and “explain”) and three corresponding components for their writing: (a) identify an individual discussed recently, (b) summarize the individual’s work, and (c) highlight the impact of that work.

The second stage, make a Plan, teaches students to outline their response to the writing prompt (Bashir & Singer, 2006; Singer & Bashir, 2004). Students determine their goals for the writing, select a topic, and generate a checklist of tasks to accomplish based on the action words and number of parts the prompt requires. Students also select an appropriate graphic organizer to scaffold their idea generation and the organization process (e.g., webs, Venn diagrams, timelines); for example, a student writing about Napoleon Bonaparte may select a timeline to chart his accomplishments in relation to other relevant historical events. As a caveat, pre-made

organizers provide a set structure that may not meet the individual needs of every student. Thus, it may be beneficial to train students to generate their own graphic organizers to ensure their ideas will not be limited by the design of pre-made organizers. At the end of the make a Plan stage, students reflect on whether they have generated enough ideas to begin mapping them using the graphic organizer or if they are missing any relevant information.

In the Organize stage of EmPOWER, students map their ideas using the selected graphic organizer and determine an appropriate sequence for presenting the information in their writing (e.g., chronological reporting of events, prioritization of topic importance; Bashir & Singer, 2006; Singer & Bashir, 2004). Students refer back to the writing prompt and their work in the previous stages to ensure they are generating relevant and sufficient ideas to achieve their writing goals. Thus, students are prompted to assess their idea generation, content organization, and whether more or less information is required. In the case of the example student writing about Napoleon Bonaparte, the student may have generated ideas and information about Bonaparte's exile during the make a Plan stage. However, when cued to refer back to the writing prompt in the Organize stage, the student may now deem the information irrelevant and discard it.

In the Work stage, students follow a basic template to transform the ideas generated in the previous stages into a connected writing (Bashir & Singer, 2006; Singer & Bashir, 2004). First, students begin composing the introductory paragraph by generating a topic sentence and identifying the associated subtopics. This paragraph clearly states the focus of the writing and previews important information presented later. In the following paragraphs, students provide details and supporting information (e.g., examples, facts) for each subtopic. The initial composition of these paragraphs may be brief (e.g., one or two sentences); however, students continue to finetune the writing by expanding on the information. Finally, the students compose

the conclusion to the writing using the PSST strategy (i.e., paraphrase the topic sentence, summarize the subtopics, and leave the reader thinking or feeling).

Students then return to the Evaluate stage of EmPOWER to assess the writing they have generated using the COLA checklist (i.e., Content, Organization, Language, and Appearance; Bashir & Singer, 2006; Singer & Bashir, 2004). This checklist prompts students to determine whether their writing is relevant to the prompt, well-organized, repetitive, or contains errors (e.g., spelling, grammar). In this stage, students are also encouraged to get feedback on their writing from others (e.g., teacher, peers, parents). Students then use this feedback and the COLA checklist to Re-work their writing in the last stage of EmPOWER.

Assessing EmPOWER Using a Cognitive Model of Writing

EmPOWER was designed primarily to support the language and executive functioning mechanisms that drive the writing process in school-age children with language learning disabilities and/or executive function disorders (Bashir & Singer, 2006; Singer & Bashir, 2004). As a result, this intervention supports many of the elements included in the various cognitive models of the writing process (e.g., Graham, 2018; Hayes & Flower, 1980; Scardamalia & Bereiter, 1987). In particular, Graham's (2018) model, which represents the most recent conceptualization of the cognitive components driving the writing process, identifies multiple mechanisms that EmPOWER was designed to support. Table 1 provides an overview of the alignment between EmPOWER and the components of Graham's (2018) model.

Graham's (2018) model identifies four primary facets (i.e., long-term memory resources, control mechanisms, production processes, and modulators) of the writing process as well as multiple supporting factors for each facet. The long-term memory facet and its subcomponents (i.e., knowledge and beliefs) address the writer's beliefs and knowledge of writing, his/her role

as a writer, text in general, language, and other concepts as they relate to the writer's particular circumstances as well as the type and topic of writing. The control mechanisms facet and its subcomponents (i.e., working memory, attention, and several executive control factors) focus on the implementation and coordination of multiple executive functions to successfully generate a written message. The model's production processes and its subcomponents (i.e., conceptualization, ideation, translation, transcription, and reconceptualization) address the generation of ideas and the transformation of those ideas into written language. The final facet, modulators, includes factors that may impact the writing process, such as the writer's emotional state, personality, and physiological status.

Several elements of EmPOWER's structure align with Graham's (2018) cognitive model of the writing process. All of Graham's (2018) control mechanisms are addressed by EmPOWER through the use of organizers, checklists, and goal setting strategies. These strategies support students' attention to the writing task, comprehension of the instructions for completing the task, and working memory to track both the instructions and the ideas generated in response to the prompt (Bashir & Singer, 2006; Singer & Bashir, 2004). EmPOWER's make a Plan stage directly supports Graham's (2018) planning factor for writing through direct cuing and additional supports such as templates and graphic organizers (Bashir & Singer, 2006; Singer & Bashir, 2004). Similar to the monitoring and react components of Graham's (2018) model, students using the EmPOWER approach monitor their writing to some degree throughout their engagement in all six stages, but especially during the second Evaluate stage (e.g., referring back to their writing goals, revising their written work; Bashir & Singer, 2006; Singer & Bashir, 2004).

EmPOWER addresses Graham's (2018) production processes in a similar manner. The conceptualization component (Graham, 2018) for comprehending the writing task is captured by EmPOWER's first Evaluate stage, which tasks students with analyzing the writing prompt (Bashir & Singer, 2006; Singer & Bashir, 2004). EmPOWER's make a Plan phase aligns with Graham's (2018) ideation component by directing students to plot their approach to writing through a series of pre-writing tasks (e.g., selecting a graphic organizer for generating ideas, completing a template; Bashir & Singer, 2006; Singer & Bashir, 2004). Translation and transcription (Graham, 2018) are addressed through a combination of EmPOWER's make a Plan, Organize, and Work phases (Bashir & Singer, 2006; Singer & Bashir, 2004). Students' ideas are generated and initially transformed into written language as brief notes or sentences in an organizer and then expanded upon as part of a template. Students further transform their ideas into more formal and connected writings (Graham, 2018) during the Work stage. The writing is then analyzed and, if necessary, reconceptualized during the last stages of EmPOWER (i.e., the second Evaluate stage and Re-work phase; Bashir & Singer, 2006; Singer & Bashir, 2004).

The long-term memory resources discussed by Graham (2018) are also addressed in part by EmPOWER. The first Evaluate stage specifically teaches students strategies for analyzing the writing prompt and the later stages train students to evaluate their own writing by reading what they have generated and determining if adjustments are needed (Bashir & Singer, 2006; Singer & Bashir, 2004). Thus, these aspects of EmPOWER address several factors of the reading components of knowledge described by Graham (2018). Likewise, EmPOWER supports several elements of writing knowledge (Graham, 2018) by training students to proceed through a series of steps (e.g., plan, organize, use templates and graphic organizers) that support the generation of

incrementally longer sentences for their final written product (Bashir & Singer, 2006; Singer & Bashir, 2004).

There are aspects of Graham's (2018) cognitive model for writing that are only partially supported or are not supported at all by EmPOWER (Bashir & Singer, 2006; Singer & Bashir, 2004). Specifically reading and writing knowledge (Graham, 2018) are partially supported by EmPOWER; knowledge about reading sources for relevant information, transcription skills (e.g., spelling, keyboarding, handwriting), and the intended audience for the writing are not addressed. EmPOWER also does not address other aspects of knowledge (e.g., oral language knowledge, multiple language knowledge, and listening language knowledge), the belief component of the long-term memory facet, or the potential modulators for writing (Graham, 2018).

Overall, the EmPOWER intervention addresses many of the components identified in Graham's (2018) cognitive model; however, key components are not addressed or are only partially accounted for (Bashir & Singer, 2006; Singer & Bashir, 2004). The model's control mechanisms and production processes are fully supported by a variety of EmPOWER's stages and strategies but the knowledge subcomponent of Graham's (2018) long-term memory resources is only partially addressed. Thus, EmPOWER provides a great deal of support for the executive functioning mechanisms that underlie the writing process but only partial support for writing's language components (Bashir & Singer, 2006; Singer & Bashir, 2004). EmPOWER does not address any of the modulators or the belief subcomponent of the long-term memory resources described by Graham (2018), which further supports the conclusion that EmPOWER is an executive functioning-based intervention that also supports some language components. However, many of these components can be embedded during direct instruction to meet the

needs of individual students and EmPOWER can be further adapted for use with writers of all ages and abilities.

Adapting EmPOWER for People with Aphasia

EmPOWER was originally designed to support the expository writing endeavors of school-age children with language learning and/or executive functioning disabilities (Bashir & Singer, 2006; Singer & Bashir, 2004). However, this approach can be modified to support several genres of writing (e.g., narrative, persuasive) as well as individuals older than school-age children who may have similar language and executive functioning difficulties. For instance, people with aphasia who experience language-based writing difficulties may benefit from an adaptation of the EmPOWER approach.

Several modifications are required for the EmPOWER approach to better address the needs of people with aphasia. Because aphasia is a language-specific disorder, executive functioning remains relatively intact.. Hence, implementing the EmPOWER approach with this population necessitates a greater emphasis on including supports for language. Adapting the reading level of text provided to people with aphasia to include simple sentences that are less taxing to decode and comprehend is one such support. Likewise, including meaningful pictures in addition to text can further support comprehension of the writing task. Training people with aphasia to implement strategies to support word finding difficulties (e.g., semantic feature analysis, circumlocution) provide people with aphasia with a means of self-talk to support language-related difficulties that may impact written expression. These language supports are adaptations that will increase the potency of implementing EmPOWER with people with aphasia.

Incorporating training in the use of technology to write would also strengthen EmPOWER and is a well-documented recommendation for writing instruction (e.g., Graham et al., 2016; Graham et al., 2012; Graham, Harris, & Chambers, 2016). Using a word processing program can support the production process of writing by clearly displaying the text, thus eliminating concerns about the legibility of the handwriting while simultaneously making the written message easier to edit. Additionally, most word processors provide features such as spelling and grammar checkers, a thesaurus, and text-to-speech capabilities that further support the writing process. Spelling and grammar checkers highlight writing units to indicate errors such as misspellings, incorrect word or punctuation choice, and repetitions of the previous word that people with aphasia may not attend to when writing by hand. Access to the word processor's thesaurus and an online dictionary can combat word choice and word retrieval difficulties by allowing people with aphasia to (a) verify word meanings and (b) look up words similar to the word they are unable to retrieve and peruse synonyms for the intended word. The text-to-speech feature would allow people with aphasia to listen to their written message aloud and potentially recognize errors in the message that they would be unable to recognize when reading their messages. Such use of the text-to-speech feature would also allow people with aphasia to identify errors that the spelling and grammar checker failed to identify.

The technological supports described above are readily available as part of most word processing programs or access to the Internet. Less readily available programs, such as word prediction software and speech-to-text features, can also support the language difficulties of people with aphasia. Word prediction programs provide a list of suggested words to continue the writing process after it is initiated, which may reduce the cognitive load associated with idea generation and word retrieval during online message composition. Speech-to-text programs

allow an individual to dictate their intended message, which may reduce the cognitive load associated with dividing attention between composing and transcribing the message.

A final adaptation of EmPOWER for use with people with aphasia is to focus on functional writing tasks. EmPOWER was designed to support the expository writing of school-age children in response to a specific writing prompt. As people with aphasia tend to be adults, this form of writing is not the most functional or appropriate focus for a writing intervention aimed at this population. Instead, EmPOWER can be adapted for use with composing emails, text messages, letters, or work documentation as appropriate to each individual. These types of writings do not typically have explicitly stated goals or directions for the writing, which would necessitate the adjustment of EmPOWER's first stage (i.e., evaluate the writing prompt) to include strategies for determining the implied purpose of the writing (e.g., answering questions posed in an email or text from an acquaintance, filling out work reports on a specific topic). However, the bulk of the EmPOWER approach would not require adjustment beyond streamlining the process as individuals increase in their level of mastery for the six stages to transform the original lengthy EmPOWER approach into a more efficient process.

For example, a person with aphasia who doesn't struggle with planning and organizing his/her writing may not need to include these stages in his/her writing process. Thus, instead of following the EmPOWER approach step-by-step, this individual would be able to shorten the process to include the initial evaluate stage, work stage, second evaluate stage, and the re-work stage. This simplified version of EmPOWER would provide the person with aphasia with the support and strategies necessary to compensate for language impairments affecting comprehension and expression skills but would not require the person to engage in the Planning and Organizing stages.

Conclusion

EmPOWER is a self-regulation intervention aimed at improving the expository writing of school-age children with language learning and executive functioning disabilities (Bashir & Singer, 2006; Singer & Bashir, 2004). This intervention focuses heavily on supporting the executive functioning mechanisms driving the writing process through scaffolded instruction in the application of six writing stages (i.e., Evaluate, make a Plan, Organize, Work, Evaluate, and Re-work) and multiple supporting strategies (e.g., using a graphic organizer, expanding on a template). EmPOWER aligns with many of the cognitive writing mechanisms (i.e., control mechanisms, production processes, and certain knowledge-based long-term memory resources) identified by the most recent conceptualization of the writing process (Graham, 2018), which makes EmPOWER an appealing and theory-guided writing intervention. This analysis specifically focused on how EmPOWER can be adapted for use with people with aphasia; however, EmPOWER can be similarly modified by SLPs and teachers for use with a wide range of ages and abilities.

There is a degree of misalignment between EmPOWER and Graham's model (2018), however, that requires exploration. First, Graham's (2018) model represents the writing process in its cognitive entirety. It contains all of the cognitive components that direct and influence a writer, but these components exert differing levels of influence on writers based on factors such as age, experience, and other individual differences. Second, as EmPOWER was initially designed for use with developing writers, aspects of the misalignment to Graham's (2018) model may be attributed to differing perspectives between a model for writers in general compared to an intervention targeting school-age children.

Third, differing perspectives between fields of study may also contribute to the misalignment between EmPOWER and Graham's (2018) model in that the field of speech-language pathology views language as a more complex and nuanced skill involved in written expression than researchers in the education and special education fields. Graham's (2018) model, along with many of the other writing models developed by education and special education researchers, focuses primarily on cognitive skills and devotes less attention to linguistic skills. Writing models developed by SLPs that explore the more detailed role language plays in written expression do exist (e.g., Bashir & Singer, 2006), however, these models are less well-known than the models developed by researchers in the education and special education fields. Thus, future research on writing may benefit from the collaboration of researchers in the fields of education, special education, and speech-language pathology to develop a writing model that includes both cognitive and linguistic skills to better guide writing instruction and intervention.

Acknowledgements

The author wishes to acknowledge the supportive feedback from Dr. Karen Hux and Dr. Michael Hebert in the composition of this manuscript.

References

- Bashir, A. S. & Singer, B. D. (2006). Assisting students in becoming self-regulated writers. In T. Ukrainetz (Ed.), *Contextualized language intervention: Scaffolding PreK – 12 literacy achievement* (pp. 565-598). Greenville, SC: Thinking Publications.
- Dinnes, C. & Hux, K. (2018). A multicomponent writing intervention for a college student with mild brain injury. *Communication Disorders Quarterly*, 39(4), 490-500. doi.org/10.1177/1525740117716416
- Dinnes, C., Hux, K., Holmen, M., Martens, A., & Smith, M. (2018). *Writing changes and perceptions after traumatic brain injury: “Oh, by the way, I can’t write.”* Manuscript submitted for publication.
- Englert, C. S., Raphael, T. E., Anderson, L. M., Anthony, H. M., Fear, K. L., & Gregg, S. L. (1988). A case for writing intervention: Strategies for writing informational text. *Learning Disabilities Focus*, 3(2), 98-113.
- Graham, S. (2018). A writer(s) within community model of writing. In C. Bazerman, V. Berninger, D. Brandt, S. Graham, J. Langer, S. Murphy, P. Matsuda, D. Rowe, & M. Schleppegrell, (Eds.), *The lifespan development of writing*. Urbana, IL: National Council of English.
- Graham, S., & Perin, D. (2007a). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99, 445-476. doi: 10.1037/0022-0663.99.3.445
- Graham, S., & Perin, D. (2007b). *Writing next: Effective strategies to improve writing*

of adolescents in middle and high schools. New York, New York: Carnegie Corporation.

Graham, S., Bollinger, A., Booth Olson, C., D'Aoust, C., MacArthur, C., McCutchen, D., & Olinghouse, N. (2012). Teaching elementary school students to be effective writers: A practice guide (NCEE 2012- 4058). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from:

https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/writing_pg_062612.pdf

Graham, S., Bruch, J., Fitzgerald, J., Friedrich, L., Furgeson, J., Greene, K., Kim, J., Lyskawa, J., Olson, C.B., & Smither Wulsin, C. (2016). Teaching secondary students to write effectively (NCEE 2017-4002). Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. Retrieved from the NCEE website:

https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/wwc_secondary_writing_110116.pdf

Graham, S., Harris, K. R., & Chambers, A. B. (2016). Evidence-based practice and writing instruction: A review of reviews. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research, 2nd edition* (pp. 211-226). New York: Guilford Press.

Graham, S., Harris, K. R., & Santangelo, T. (2015). Research-based writing practices and the common core: Meta-analysis and meta-synthesis. *The Elementary School Journal*, 115(4), 498-522.

- Harris, K. R., Graham, S., & Mason, L. H. (2003). Self-regulated strategy development in the classroom: Part of a balanced approach to writing instruction for students with disabilities. *Focus on Exceptional Children*, 35(7), 1.
- Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3-30). Hillsdale NJ: Erlbaum & Associates.
- Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 57-71). Mahwah, NJ: Erlbaum.
- Ledbetter, A. K., Sohlberg, M. M., Fickas, S. F., Horney, M. A., & McIntosh, K. (2017). Evaluation of a computer-based prompting intervention to improve essay writing in undergraduates with cognitive impairment after acquired brain injury. *Neuropsychological Rehabilitation*, 1-30.
- Mason, L. H., & Graham, S. (2008). Writing instruction for adolescents with learning disabilities: Programs of intervention research. *Learning Disabilities Research & Practice*, 23, 103-112.
- National Commission on Writing (April, 2003). *The neglected "R": The need for a writing revolution*. New York: College Board. Retrieved from http://www.collegeboard.com/prod_downloads/writingcom/neglectedr.pdf
- National Commission on Writing (May, 2006). *Writing and school reform*. New York: College Board. Retrieved from

http://www.collegeboard.com/prod_downloads/writingcom/writing-school-reform-natl-comm-writing.pdf

- Scardamalia, M., & Bereiter, C. (1987). Knowledge telling and knowledge transforming in written composition. In S. Rosenberg (Ed.), *Advances in applied psycholinguistics: Reading, writing, and language learning* (vol. 2, pp. 142-175). New York: Cambridge University Press.
- Singer, B. D. & Bashir, A. S. (2004). EmPOWER: A strategy for teaching students with language learning disabilities how to write expository text. In E. R. Silliman & L. Wilkinson (Eds.), *Language and literacy learning* (pp. 239-272). New York: Guilford.
- Zimmerman, B., & Reisemberg, R. (1997). Becoming a self-regulated writer: A social cognitive perspective. *Contemporary Educational Psychology*, 22, 73-101.

Table 1. Comparison of EmPOWER's Components to Graham's (2018) Cognitive Model

Components of Graham's (2018) cognitive model	EmPOWER
Control mechanisms	
Attention	✓
Working memory	✓
Executive control	✓
Long-term memory resources	
Knowledge	/
Beliefs	-
Production processes	
Conceptualization	✓
Ideation	✓
Translation	✓
Transcription	✓
Reconceptualization	✓
Modulators	
Emotions	-
Physical state	-
Personality traits	-

Note. ✓ = component is fully addressed by EmPOWER; / = aspects of the component is addressed by EmPOWER- = component is addressed by EmPOWER.

Learning to Lesson Plan: A Mentor's Impact on Pre-service Teachers

Kelly M. Gomez-Johnson

Department of Teaching, Learning, and Teacher Education

University of Nebraska-Lincoln

Abstract

Lesson planning is considered an essential skill of teachers. As pre-service teachers first encounter the fundamental principles of planning for instruction, the complexity of planning to support the rigorous learning goals of content, curriculum, and individual student needs could be daunting. The mixed methods study explored how mentoring influenced early-program pre-service teachers' self-efficacy (Bandura, 1997) and progression through stages of concerns (Fuller, 1969) in relation to lesson planning. Participants, secondary early-program pre-service teachers enrolled in a Midwestern teacher preparation program, included a target group who received mentoring and a comparison group who did not. Using constant comparison techniques guided by the theoretical frameworks, researchers gathered and analyzed qualitative target group data from participant reflections and mentor conference memos. Researchers collected and analyzed quantitative data using Likert-type survey questions also linked to the frameworks of this study. Integrated findings from the qualitative and quantitative data revealed that when a mentor provided metacognitive modeling, addressed anxiety, recognized incremental victories, and offered focused feedback, pre-service teachers' concerns were addressed and their efficacy increased related to lesson planning.

Keywords: empirical research, lesson planning, pre-service teacher education, mixed methods

doi: 10.32873/unl.dc.ne002

Many professions allow for a substantial learning curve at the outset of an individual's career. However, the field of education expects a teacher, immediately after completing a preparation program, to independently and effectively manage and monitor P-12 student learning. As beginning teachers start their careers under these expectations, they are likely to encounter high-stakes evaluations and concerning attrition rates. Understanding this context, teacher preparation programs should do everything possible to advance the skills of pre-service teachers as soon and as far as possible (Ingersoll, Merrill, & May, 2012).

Teacher preparation programs vary, but most address standards which include the fundamental skill of lesson planning (Chief State School Officers, 2013; Tummons, 2010). Despite being a basic teaching skill, the challenges of learning to lesson plan are well documented in the literature. In the following study, researchers addressed how intentional mentorship influences pre-service teachers enrolled in a traditional four-year university teacher preparation program. Specifically, researchers examined how mentoring impacts pre-service teachers' self-efficacy and levels of concern as they develop the most basic of teaching skills – lesson planning.

Literature Review

Pre-service Teachers and Lesson Planning

Thoughtful lesson planning is complex, and teaching pre-service teachers the implicit cognitive processes involved in lesson planning is difficult (John, 2006; Jones, Jones, & Vermette, 2011; Tummons, 2010; Rusznyak & Walton, 2011). Many teacher educators teach pre-service teachers how to create a product - a lesson plan comprised of required individual components, such as objectives, materials list, procedures, strategies, assessments, and closures (Drost & Levine, 2015). Despite the explicit teaching of individual lesson plan components,

many pre-service teachers struggle with writing and assessing objectives (Jones, et al., 2011; Eun Kyung, 2012), as well as planning effective questions, (Purdum-Cassidy, Nesmith, Meyer & Cooper, 2015), student engagement (Jones, et al., 2011), and instructional strategies (Ruys, Keer, & Aelterman, 2012). Furthermore, pre-service teachers' response to this product-based approach may be primarily one of complying with technical expectations of the required lesson plan components rather than engaging in the complex cognitive processes involved in lesson planning (Tummons, 2010).

Preparation programs' vision and selection of key lesson plan components matters because most preservice teachers adopt components (e.g. objectives, procedures, strategies) required by their program. Only later in their careers will they begin to integrate important complex variables such as differentiation, learning styles, and cross-curricular connections into their planning (Fielder, 2014; John, 2006; Johnson 2000). This evolving nature of planning presents a further challenge in that lesson planning processes taught within preparation programs may not match the planning practices of more experienced classroom teachers with whom pre-service teachers are placed for clinical experiences (Eun Kyung, 2012).

Pre-service teachers themselves recognize their struggles with lesson planning. In an extensive review of nearly 50 years of literature, pre-service teachers consistently identified components of lesson planning (Cherubini, 2009) as a concern. Unsurprisingly, planning is a greater concern for pre-service teachers in the early stages than for those in more advanced stages of their preparation programs (Cho, 2017; Wee, Weber, & Park, 2014).

Mentoring Pre-service Teachers

Mentoring could potentially address some challenges of learning to lesson plan. Mentoring is defined as one-to-one support of a novice or less experienced practitioner (mentee)

by a more experienced practitioner (mentor), designed to enhance the professional learning of the mentee and to mediate their induction into the profession (Orland-Barak, 2014; Hobson, Ashby, Malderez, & Tomlinson, 2009; Hobson, Harris, Buckner-Manley, & Smith, 2012). Mentoring can make novice teachers “more autonomous as professionals, reflective of experience, and aware of the students' needs” (Kelehear, 2003, p. 35). Although mentoring is common in teacher preparation programs, who assumes the mentor role as well as the duration, context, and purpose of mentoring varies (Ambrosetti, 2009; Ambrosetti, Knight, & Dekkers, 2014; Hobson et al., 2009; Hobson et al., 2012).

Despite variation in the structure of mentoring, researchers have identified mentor qualities particularly helpful to pre-service teachers. First, mentors should have strong instructional skills in order to help pre-service teachers address the common and difficult task of transferring theory into practice (Orland-Barak & Hasin, 2010; Hudson, Usak, & Savran-Gencer, 2009; Schneider, 2007). Second, the ability of the mentor to provide emotional support is critical to foster a positive and productive relationship with pre-service teachers (Orland-Barak & Hasin, 2010; Feiman-Nemser & Rosaen, 1997a). Mentors foster effective relationships through affinity and personal commitment to the success of the mentee as well as by establishing goals and expectations with the mentee (Feiman-Nemser & Rosaen, 1997b). Finally, trust enhances the ability of mentors to hone skills, learn from mistakes, build confidence, and reach goals (Wasonga, Wanzare, & Dawo, 2015; Bennis, 2003; Kouzes & Posner, 2002; Northouse, 2004, Schville, 1997). Effective mentors establish trust with pre-service teachers by being flexible, enthusiastic, friendly, and willing to help (Orland-Barak & Hasin, 2010; Hudson & Nguyen, 2008).

Theoretical Framework

This study, framed by the theories of Fuller (1969) and Bandura (1997), examined how mentoring support in the early stages of a teacher preparation program might resolve pre-service teachers' struggles related to learning to lesson plan. Fuller (1969) outlined three stages of concerns pre-service teachers encounter: (a) self concerns related to acceptance and self-adequacy; (b) task concerns related to the day-to-day responsibilities of teaching, such as lesson planning; and (c) impact concerns related to P-12 student learning. If unresolved, concerns could inhibit an individual's development as a teacher (Fuller, 1969).

Bandura's Social Cognitive Theory (1997) offers an efficacy framework to address pre-service teacher concerns identified by Fuller (1969) and improve their confidence. According to Bandura (1977, 1997) teacher efficacy is based on four components: (a) early mastery experiences; (b) emotional states; such as excitement or anxiety which lead to perseverance; (c) vicarious experiences; (d) actionable feedback from a mentor, coach, or respected colleague. Early support for pre-service teachers is crucial, as positive change in teacher efficacy is more likely to occur during initial stages of becoming a teacher (Hoy & Spero, 2005) and research has found a direct relationship between teacher self-efficacy and increased P-12 student performance (Ashton, 1984; Bandura, 1977, 1997; Dembo & Gibson, 1985; Denzine, Cooney, & McKenzie, 2005; Hoy & Spero, 2005; Lin & Gorrell, 2000).

Rationale and Significance

Drost and Levine (2015) called for continued research on how to support pre-service teachers in the process of planning. Others have called for teacher educators to reconsider how to support pre-service teachers not only as they learn the individual components of lesson planning but also as they develop an individualized or personal process for planning (Eun Kyung, 2012;

Peralta & Burns, 2012).

Jones, et al. (2011) suggested teacher educators “intentionally discuss the lesson planning and delivery decisions” (p. 855) with pre-service teachers at the time lesson plans are being written. However, research to inform a discursive, shared approach to teaching lesson planning is notably absent from the literature (Hobson et al., 2009). Existing studies have been limited to pre-service teachers during student teaching experiences (Morton & Gray, 2010; Smith, 2005). Furthermore, much of the existing research on mentoring pre-service teachers is within the supervisory context of student teaching and other field experiences where mentors have an evaluative as well as a mentoring role (Ambrosetti, 2009; Feiman-Nemser & Rosaen, 1997). This study explores mentoring as a means to decrease concerns and increase the efficacy of pre-service teachers in the area of lesson planning, a skill which has been identified as essential for beginning teachers (CCSS0, 2013).

Purpose Statement

The purpose of this research was to examine the impact of mentoring as it related to early-program, pre-service teachers’ concerns and efficacy related to lesson planning.

Two research questions were addressed.

1. How will mentoring address lesson planning concerns of early-program, pre-service teachers?
2. How will mentoring support early-program, pre-service teachers’ efficacy in regard to lesson planning?

Methodology

Researchers employed a sequential explanatory mixed methods design. This approach allowed the quantitative data analysis to refine and provide general understanding to the

qualitative data and overarching research problem concerning mentoring support for pre-service teachers. The researchers employed two distinct methods phases: qualitative followed by quantitative (Creswell, Plano Clark, Gutmann, & Hanson, 2003). The interpretation stage included the collection and analysis of both qualitative and quantitative data. Following this interpretation stage, qualitative and quantitative data were connected during the integration stage for further analyses. With complex design structures, visual models (see Figure 1) are considered best practice in expressing mixed methods procedures (Creswell, 2005; Tashakkori & Teddlie, 1998).

Figure 1.

Visual model for the mixed methods design structure of this study

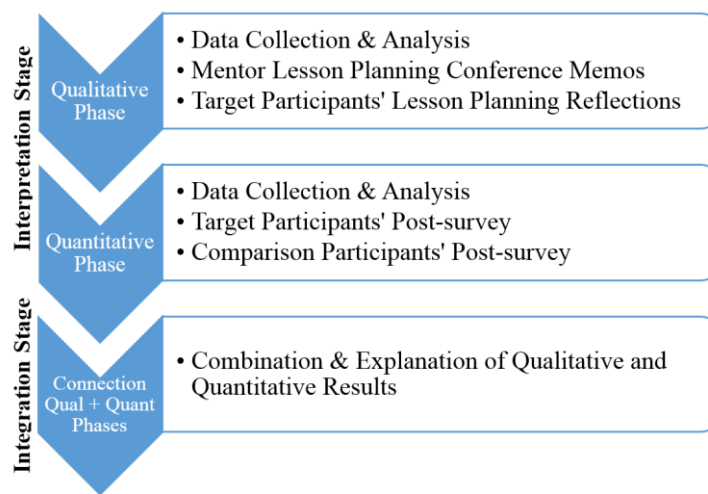


Figure 1. Visual model for the mixed methods design structure of this study.

The qualitative components of the mixed methods design utilized instrumental case study methods (Stake, 2005). The primary researcher examined the mentor's conference memos and early-program pre-service teachers' reflections. The memos and reflections were collected within a finite time frame, the specific context of a teacher preparation course, and from a relatively small number of pre-service teachers (N=11).

The quantitative component utilized survey methods. Two groups of pre-service teachers, a target (N=10) and comparison (N=21) group, completed a survey. The research team compiled and compared data from the groups using descriptive statistics and independent samples t-test analysis. The research team deemed the t-test to be an appropriate statistical analysis given the sample size (De Winter, 2013).

The mixed methods design minimized the quantitative disadvantage of a small sample size and mitigated researcher bias in the qualitative findings (Creswell et al., 2003; Greene & Caracelli, 1997). The research team gave equal priority to the data, analyzing and integrating both quantitative and qualitative data through the lenses of Fuller's stages of concern and Bandura's concept of teacher efficacy.

Participants

Researchers used a convenience sample of early-program pre-service teachers recruited from a first-year lesson planning course in an undergraduate, secondary teacher preparation program at a mid-sized Midwestern university in the United States. Students enrolled in the course represented varied secondary education subject areas. Participants in the target (11) and comparison (21) groups were drawn from two distinct sections of the course and taught by the same instructor who was also one of the researchers. Both sections included the same assignments and grading criteria, as well as a required 60-hour practicum experience. Participation in the study was voluntary. Participants signed informed consent prior to taking part in the study.

The mentor, also the primary researcher, was a graduate assistant with 13 years of teaching experience who did not evaluate or grade any participants or other students enrolled in the course or clinical experience. An internal graduate research grant funded the research, and a

university institutional review board approved the research.

Data Collection

Data collection spanned a 16-week semester. As a requirement of the course, participants in the target and comparison groups completed four lesson plans in their designated content areas. During the first 10 weeks, participants created the initial three plans for a hypothetical class of secondary students. Participants spent the remaining six weeks of the semester in a 60-hour practicum experience in a secondary school. For the fourth lesson plan, participants planned and taught a lesson as stipulated by a classroom teacher in the practicum setting. Data were collected from three sources: (a) mentor conference observation memos; (b) participant reflections; (c) end-of-course surveys.

Mentor Conference Observations Memos

In addition to the lesson plan assignments, members of the target group participated in three, 15-20 minute mentor conferences scheduled one week prior to the submission of each of the first three lesson plans. During the conferences, the mentor asked participants to identify areas of strength and points of concern related to their lesson plans. Answers to these questions guided the discussion for the remainder of the conference. The mentor offered in-person or telephone conference options for the second and third conference as time constraints became a factor for some target group participants. Approximately one-third of the target group took advantage of the telephone option for one or both conferences. Following each conference, the mentor wrote observational memos describing the lesson plan, participant concerns, and the support and metacognitive modeling offered by the mentor.

Participant Reflections

Within two weeks after completing the first three lesson plans, target group participants

wrote an initial reflection regarding their experiences. In their reflections, participants answered seven questions related to their lesson planning experiences (Figure 2). To encourage honest responses, participants submitted reflections using an anonymous online response system. A second reflection, also anonymous, focused on the final plan, which participants created without the mentor assistance. Participants completed the second reflection after teaching the final plan during their practicum experience. The second reflection targeted the identification of unplanned or surprising moments during the lesson (Figure 2).

Figure 2.

Participant reflection questions

<u>Initial Reflection Questions</u>	
1.	What about the lesson planning process concerned you the most?
2.	Did you notice a change in concerns over the course of planning sessions 1-3? For example, were you uncertain or apprehensive about aspects of lesson 1 that no longer worried you in lesson 2 or 3? Please explain.
3.	Over the course of planning lessons 1-3, what did you feel most comfortable with?
4.	Did you start the class at this level of comfort, or did it grow over time? Please explain.
5.	Were there any internal or external reasons you can pinpoint for helping you to become a more effective lesson planner?
6.	Were there any internal or external reasons you can pinpoint that hindered you from becoming a more efficient lesson planner?
7.	Optional: Is there anything else you would like the researcher to know about your experience in this study? Feel free to reflect on anything not covered in the questions above.
<u>Second Reflection Questions</u>	
1.	What about teaching the fourth lesson went differently than planned? How did you respond to the unplanned moments?
2.	What surprised you about the planning and teaching of this lesson?
3.	How has your lesson planning evolved throughout this course experience?

End-of-course survey. The four-point Likert-type survey addressed lesson planning

through the lens of Fuller's stages of concern. Items 1-5 addressed self concerns, items 6-10 task concerns, and items 11-15 impact concerns. The primary researcher's literature review of Fuller's stages of concern formed the basis of the survey and paralleled a lengthier validated survey related to Fuller's stages of concern for student teachers (Author, 2003). The course instructor as well as a program faculty member, both with previous experience teaching the course, reviewed the survey and provided suggestions to the primary researcher. Ten of 11 target group and 21 comparison group participants completed the survey using an anonymous online survey tool.

Data Analysis

Following completion of the course, participants' reflections and mentor's observation memos were investigated a posteriori (Constas, 1992) for recurring categories and themes using the three-stage constant comparison model outlined by Strauss & Corbin (1998). In the open coding stage, the primary researcher read reflections multiple times, highlighting statements that characterized and defined participant responses and/or observations of the mentor. During this stage, the theoretical frameworks provided an initial reference for themes (Constas, 1992). Once the primary researcher found no new defining codes, members of the research team organized the initial codes into axial codes representing broader categories participants expressed or the mentor had observed in the conferences. In the final stage, researchers integrated axial codes and assigned names to the themes (Constas, 1992). The primary researcher used tree diagrams (Creswell, 1998) to visually analyze the: (a) relationship between mentor observation memos data and the theoretical frameworks; (b) evolution of concerns and efficacy evidenced in initial reflections of participants; (c) themes from the second reflections of participants.

The quantitative analysis included calculating means and standard deviations for survey

items and using an independent samples t-test to compare the mean scores of the target and comparison groups' perceptions of concerns related to lesson planning. Researchers chose independent samples t-test to examine potential differences in the groups. This test assumes the difference in mean of the dependent variable (perceptions of concern) is found because of the influence of the independent variable (mentoring support). All analyses set 95% as the confidence interval, and $p < .05$ was considered statistically significant. Researchers measured the internal consistency of the 15-item survey using the Cronbach's Alpha reliability statistic. A coefficient of .910 indicated a high level of internal consistency for the scale given the sample. The researchers analyzed survey data using SPSS, version 11.

Findings-Qualitative Analysis

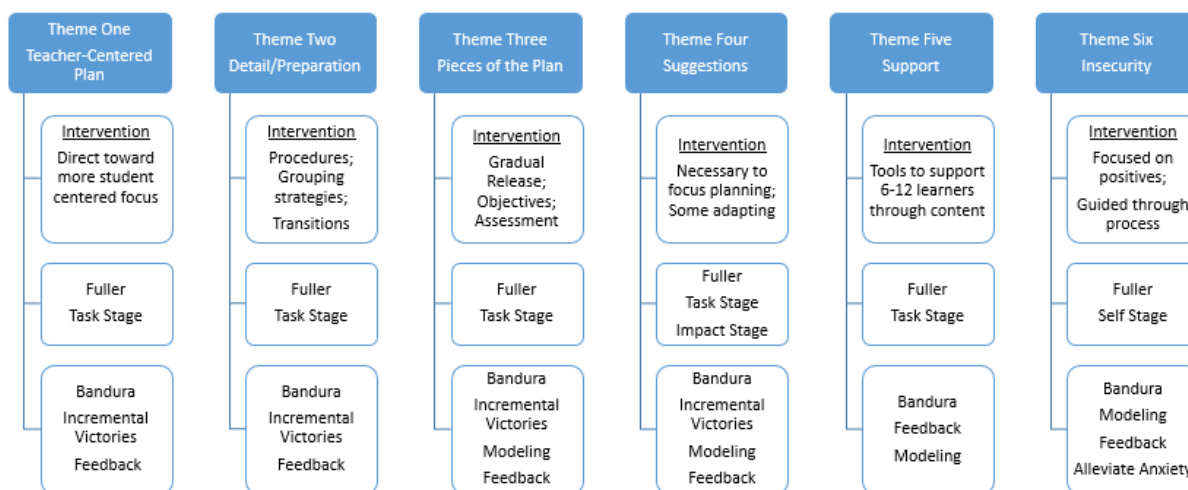
Mentor Observations After Conference One

Six themes, shown in Figure 3, emerged from the first set of mentor observation notes. Theme 1 represented participants' over-reliance on teacher-centered direct instruction as the primary method of delivery. Therefore, the mentor interventions focused on directing participants toward more student-centered methods. Theme 2 signified a lack of detail in the plans. The mentor reinforced the need to include steps for procedures, intentional grouping, and transitions where appropriate. Theme 3 reflected participants' struggle with individual components of the plan. Intervention included reviewing delivery models, re-working objectives, and adding formative assessments to the plans. Theme 4 centered on mentor suggestions to focus the direction of the lesson plan as well as ideas for aligned activities. All participants were receptive to feedback and guidance. However, many added the mentor's suggestions verbatim into their plans. Few participants expanded or elaborated on mentor suggestions. Additional mentor interventions included modeling strategies to deepen participant understanding of

concepts. Theme 5 emphasized participants' lack of instructional techniques to guide secondary students through subject-specific content. The mentor interventions helped participants plan for strategies such as note-taking, graphic organizers, and text annotation. Finally, Theme 6 highlighted participants' insecurity related to the lesson planning process. Interventions focused on identifying strengths within plans and recognizing successful completion of individual components of a lesson plan as important initial steps toward completion of the entire plan.

Figure 3

Self-efficacy for the task at hand- mentor reflections from conference one



The mentor observation memos from three participants provide evidence of the themes.

“A lot of Alex’s lesson was teacher centered. We worked together to break up the lesson so it became more student centered. She used suggestions to do the first third whole class, the second third small group and the final third independently. Inserting student engagement strategies as well as creative grouping was necessary as those were details she didn’t yet understand.”

Themes One, Four, and Five- Teacher-Centered Plan, Suggestions, and Support

“I basically combed through the lesson [plan] to tweak it for missing procedural aspects. She (Ashley) was open to any suggestions that would make her plan better. She was very intent

on getting all points possible... I talked her through questions like: Do they cut the shapes out? Do they draw the shapes? Can they use the shapes more than once? Can they cut a circle in half to make a dome?”

Themes Two, Four, and Six- Detail/Preparation, Suggestions, and Insecurity

“She (Terry) started out unsure of how to write a clear and measurable objective and had a plan that was lacking in detail, instructional strategies and intentional grouping. She was unclear on the lesson planning process and our first conference was spent going over every section in detail.”

Theme Two and Three- Detail/Preparation, Pieces of the Plan

“Analysis of the observation notes revealed evidence of Fuller’s stages of concern. Participants’ anxiety associated with successful completion (i.e., getting a satisfactory grade) of the lesson plan was indicative of self concerns. Participants also demonstrated task concerns in their desire to perform an important skill, lesson planning, with fidelity. Impact concerns emerged as participants began to integrate mentor instructional suggestions to enhance student learning opportunities into their plans.”

Examination of the observation memos also revealed interventions aligned to Bandura’s sources of efficacy attainment. The interventions of the mentor fell into four categories: (a) identifying and praising incremental victories (glimpses of participants’ early mastery of planning); (b) alleviating participant anxiety; (c) creating vicarious experiences through the mentor’s metacognitive modeling; (d) providing focused feedback.

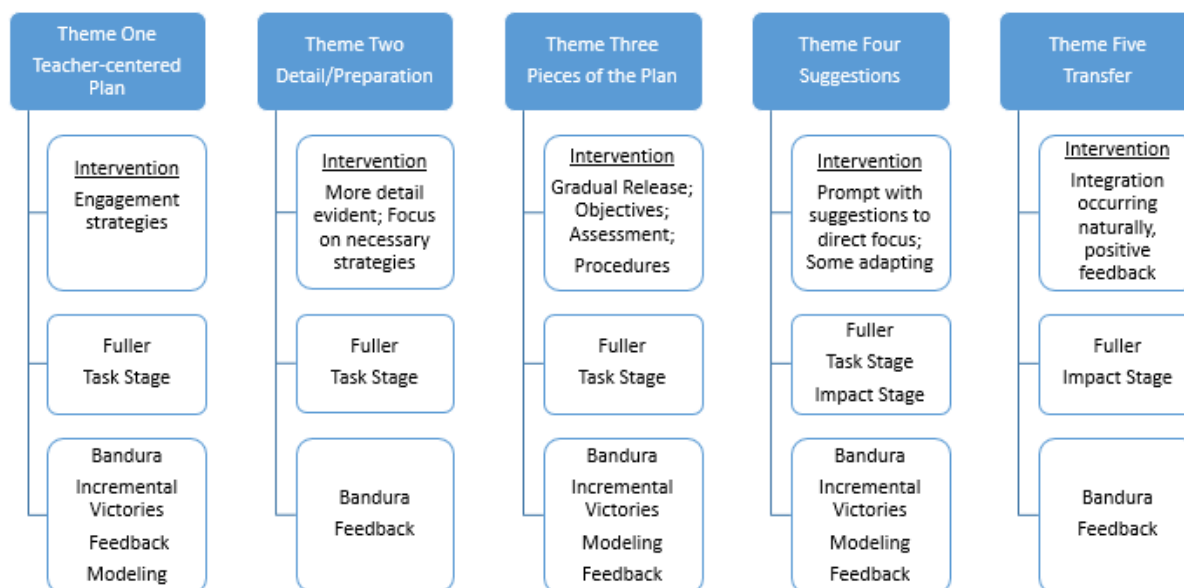
Mentor Observations After Conference Two

Figure 4 depicts five themes, stages of concern, and interventions identified within the second set of observation memos. Themes 1-4 paralleled those following the first conference

although participants' progress was noted in each area. A new theme reflecting participants' ability to transfer and integrate suggestions from the first conference into subsequent lesson plans was noted. The need for support and the insecurity of participants had subsided by the time of the second conference.

Figure 4.

Self-efficacy to shift thinking- mentor reflections from conference two



Quotes taken from the same three participants noted above, exemplify the themes and a progression from first to second conference.

She (Alex) grouped more intentionally this time on her own...Students were to identify the story elements. Her mode of questioning needed developing in that she wanted to just have the teacher ask questions of the whole class. Then she took my suggestion to use a picture book as a tool to read a complete short story that would be easier to manage for the purposes of her objective – to map the 5 elements of a short story.

Themes One and Five- Teacher-centered, Transfer

“Ashley was very intent on doing well in this class and was not afraid to add or subtract

where necessary to make this happen.”

Themes Two and Four- Details, Suggestions

“I could tell Terry was more confident with the process in that she was more prepared with the amount of work she put into this lesson. She was missing some of the procedural components so important for new teachers to understand, but she was so open and eager to include versions of my suggestions, or her own to make that happen.”

Themes Three and Four- Pieces of the Plan, Suggestions

“Fuller’s stages were now primarily task focused, with the impact stage emerging in Themes Four and Five. Interventions related to Bandura’s theory focused primarily on helping participants achieve incremental victories within the second plan, and creating opportunities for vicarious experiences through modeling the metacognitive process involved in planning, and providing focused feedback.”

Mentor Observations After Conference Three

Figure 5 includes the themes found in the third set of mentor observation memos as well as the stages of concern and interventions. By the end of the third lesson conference, two new themes emerged. The first new theme reflected the majority of participants who arrived at conference three with a complete plan. The mentor intervened with positive feedback. Mentor statements drawn from the observation memos from the same three participants referenced in conferences one and two illustrate the first new theme.

Alex came well prepared with a lesson that had more detail than previous lessons with special attention to including student engagement and planned grouping. Her activities paced well and learning was not all teacher directed.

Theme One- Comprehensive Plan

“I would say by the third lesson she (Ashely) was more intent on doing whatever it took to arrive at a good product instead of so much concern about her grade.”

Theme One- Comprehensive Plan

“She (Terry) just kept on growing. This time she added the following: interactive AS [anticipatory set] with color coding to be connected to later learning; formative assessment throughout (in italics); intentional grouping; purposeful observation; and a graphic organizer for the video portion.... I believe Terry had begun to distinguish herself as a learner who was evolving to more than just checking the assignment box.”

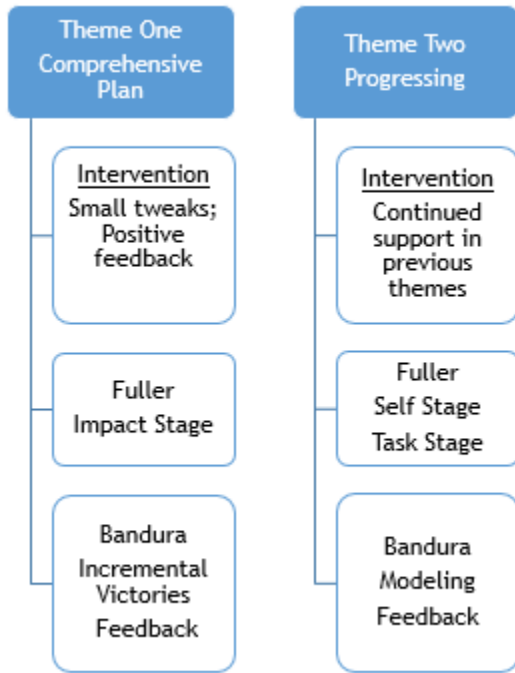
Theme One- Comprehensive Plan

The second new theme represented two participants who did not have complete plans. Both of the participants continued to be unable to produce a coherent lesson plan and needed extensive support. However, neither implemented the mentor’s suggestions.

“I worked with him (Leon) explaining/suggesting turning his simple story into a puzzle (cut sentences into strips) and put them in envelopes. Then, pass them to small groups and have them try to put the story in order...I told him this was only one way to engage learners and he could come up with his own idea. He didn’t do either. He left the written, skeletal text he arrived with exactly as he had it.”

Theme Two- Progressing

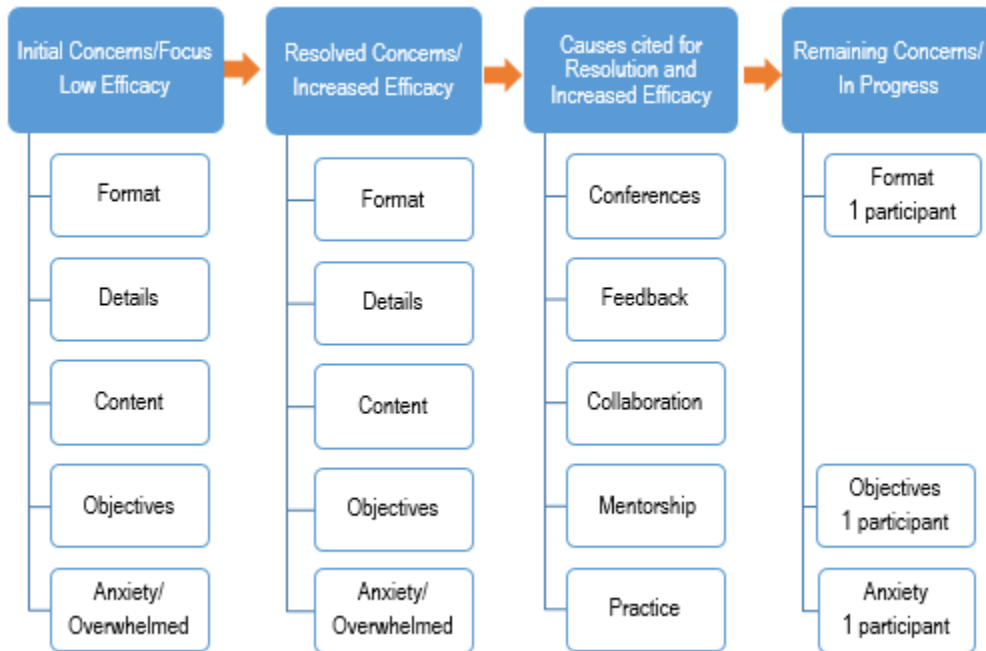
“I could tell it was just the start of a lesson because it was a very broad outline that I could understand only after talking with her (Julie) on the phone ...I encouraged her to include an instructional strategy... However, she did not do this in her submitted lesson.”

Figure 5*Self-efficacy to impact learning – mentor reflections from conference three*

Although most participants had reached Fuller’s impact stage, two remained in the self and task stages. Their lesson plans continued to be teacher-centered, and they struggled with writing a complete plan. Interventions related to Bandura’s theory focused on recognizing incremental victories represented within the comprehensive plans, continuing the metacognitive modeling of planning, and providing focused feedback.

Participant Reflections

Participant reflections provided the second source of qualitative data. Participants wrote initial reflections after submitting the first three lesson plans and a second reflection after completing and teaching the fourth plan. Examination of the reflections uncovered common topics of concerns, ways in which concerns were resolved, and a progression toward increased efficacy represented in Figure 6.

Figure 6*Path to self-efficacy in lesson planning- participants' initial reflection*

When asked to review the lesson planning process, participants' initial concerns focused on format, details, content, and objectives. Additionally, some participants noted anxiety related to lesson planning and expressed being overwhelmed with the assignment. However, after completing the last of the three plans, participants perceived those concerns as resolved and noted a sense of confidence with writing plans. Participants' reported increased efficacy was a result of conferences with the mentor, feedback, collaboration, and practice. One participant voiced ongoing concern with lesson planning format, writing objectives, and continued anxiety.

The following comments, taken from four participants' anonymous initial reflections, exemplify the shift in Fuller's stages of concerns and increasing efficacy which occurred throughout the first three mentor conferences.

"I really had trouble with objectives when we first started planning lessons, but after

conferences with the mentor, I felt much more confident about where my lessons were going. Now, as I'm prepping lessons for practicum, I feel completely comfortable with objectives."

"Over time I became more comfortable with the idea of planning towards students. At first it was planning for an assignment, but by the end of was keeping in mind what my practicum classroom of students would be able to handle and keep up with."

"Feedback was well thought out and very helpful. The mentor had great ideas.

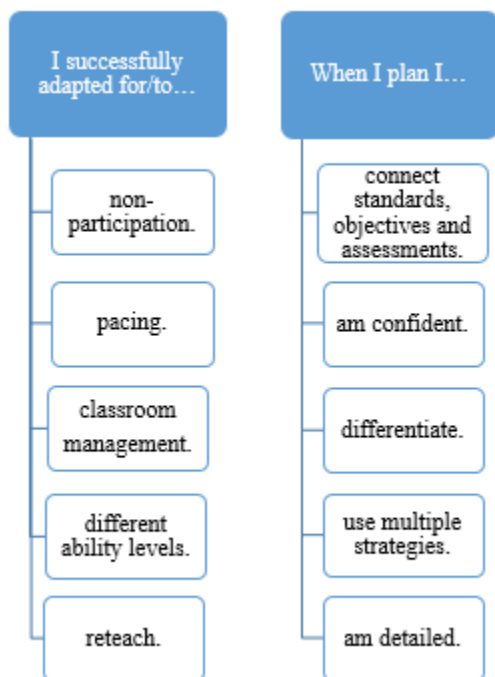
Sometimes talking to a teacher can be hard, but she made everything easy and put me at ease about what to do in my lesson. She always stayed positive."

"I felt that I internalized a lot of the mentor's advice after conferencing with her. For example, in lesson plan two she gave me a lot of ways of presenting ideas, but in lesson three I felt more confident in determining for myself how to present topics and came to the conference with a nearly complete lesson plan."

The second participant reflections occurred after having written and taught a lesson plan in their practicum. Reflections were again examined to uncover common topics shown in Figure

Figure 7

Self-efficacy transferred- participants' second reflection



After reflecting on the process of planning and teaching a lesson, most participants recognized the need for flexibility to adapt to a specific context. The following comments reveal participants' perspectives on the evolution of their lesson planning knowledge, skills, and confidence throughout the semester and under the guidance of a mentor.

"I was able to make small changes as I progressed through the lessons. As I moved onto more and more lessons I noticed myself feeling more confident in the parts of the lesson plan. I also noticed that I did not make as many errors because I had received feedback and made changes on the lessons prior."

"I am now able to think and plan for multiple scenarios. Instead of assuming that my lesson plan will go one way, I can now think ahead and decide what I will do if it doesn't. I also have been able to progressively differentiate my lesson plans."

"It (lesson planning) has gotten much more detailed and I've developed a lot more

confidence in using the Gradual Release model. I've also gotten more confident with knowing what activities and assessments I think will be best to meet the lesson objective.”

Participants stated they were now able to complete components of the plan, such as objectives, and assessments, and were confident about the planning process. Participants were aware of the importance of detail and felt capable of incorporating additional details in future lesson plans. They differentiated their instruction for varied student abilities and used multiple strategies when planning.

Findings-Quantitative Analysis

At the end of the semester, participants in both target and comparison groups completed a Likert-type survey on lesson planning. Researchers used an independent samples t-test to analyze differences between target and comparison group data. Table 1 includes the means, standard deviations, and t-scores as related to Fuller’s three stages of concern: self, task, and impact.

Table 1*Pre-service Teachers' Perceptions of Planning Related to Fuller's Concern Theory*

	Control	Experimental	<i>t</i>	<i>df</i>
	<u>Concern for Self</u>			
Question 1: My instructor gave positive feedback on my objectives.	3.52 (.51)	3.90 (.32)	2.13*	29
Question 2: I can plan anticipatory sets that will be well received by my students.	3.19 (.40)	3.90 (.32)	4.49**	29
Question 3: I can plan assessments that are appropriate and insure I am doing my job well.	3.20 (.62)	3.50 (.53)	1.32	28
Question 4: I planned questions to ask students to help me determine if I was on the right track.	3.19 (.81)	3.50 (.53)	1.09	29
Question 5: I planned closures that would help me conclude the lesson.	3.19 (.52)	3.70 (.48)	2.64*	29
	<u>Concern for Task</u>			
Question 6: My lesson objectives met the requirements of the assignment.	3.24 (.43)	3.90 (.32)	4.28**	29
Question 7: My anticipatory sets were planned according to instructor guidelines.	3.38 (.59)	3.80 (.42)	2.01	29
Question 8: I planned appropriate assessments in my lessons.	3.38 (.50)	3.70 (.48)	1.68	29
Question 9: I planned pertinent questions according to the correct levels of Bloom's Taxonomy.	3.47 (.51)	3.40 (.52)	-.39	29
Question 10: My lesson closures were planned to summarize content and manage end-of-class procedures.	3.14 (.48)	3.80 (.42)	3.71**	29
	<u>Concern for Impact</u>			
Question 11: I connected lesson objectives to the standards to insure meaningful learning.	3.23 (.62)	3.60 (.52)	1.59	29
Question 12: I planned anticipatory sets to engage learners to maximize understanding.	3.23 (.54)	3.80 (.42)	2.90*	29
Question 13: I included opportunities for assessment into my lesson plans and understand how results can be used to inform future instruction.	3.25 (.55)	3.40 (.52)	.72	28
Question 14: I planned questions to develop students' critical thinking skills.	3.52 (.60)	3.30 (.67)	-.93	29
Question 15: I planned closure techniques to review and reinforce lesson content.	3.24 (.54)	3.60 (.52)	1.78	29

Note. * = $p \leq .05$, ** = $p \leq .001$. Standard Deviations appear in parentheses below means.

Items 1-5 focused on participants' perceptions of planning related to Fuller's concern for self. Based on descriptive statistics, pre-service teachers in the target group were more confident on all five items compared to the comparison group. Within the self-concerns stage, the target group's efficacy means differed significantly from the comparison group's means in the following lesson planning categories: behavioral objectives ($t(29) = 2.13, p < .05$), anticipatory sets ($t(29) = 4.49, p < .001$), and lesson closures ($t(29) = 2.64, p < .05$).

Items 6-10 focused on perceptions of planning related to Fuller's concern for task. Pre-service teachers in the target group reported efficacy scores significantly higher than comparison group efficacy counterparts in the tasks of writing behavioral objectives ($t(29) = 4.28, p < .001$) and using closure to summarize content and manage procedures ($t(29) = 3.71, p < .001$).

Items 11-15 focused on perceptions of planning related to Fuller's concern for impact. Pre-service teacher's efficacy scores for planning anticipatory sets to impact student learning was significantly higher in the target group than the comparison group ($t(29) = 2.90, p < .05$).

Interpretation

Findings of this study described how mentoring increased efficacy and lessened concerns for pre-service teachers' lesson planning development. As in previous research (Kelehear, 2003; Jones, et al., 2011), mentoring mattered for participants in this study, both in terms of addressing their concerns and their efficacy. The data demonstrated pre-service teachers who were mentored in the area of lesson planning perceived their ability to address a number of lesson plan concerns to be significantly stronger than those who were not mentored. The pre-service teachers' concerns were addressed as their efficacy increased, and mentoring was a critical factor contributing to this self-perceived confidence. Pre-service teachers, even at this early stage in their preparation, began to include differentiation into their plans, something often not done until

a teacher has gained several years of experience.

Pre-service teachers and the mentor provided rich descriptions of the impact of mentoring. Initially, planning was based on the pre-services teachers' internal need to complete assignments to fulfill course requirements. However, after observing metacognitive modeling of the mentor, addressing anxiety, achieving incremental victories, and receiving focused feedback, the focus of pre-service teachers turned outward toward planning for student learning: Fuller's concern for impact. Mentor conferences provided timely feedback, encouragement, and lesson planning suggestions to pre-service teachers who subsequently felt more capable of planning. These interventions support literature addressing pre-services teachers' varied approaches to lesson planning assignments (Tummons, 2010) and the necessary components to increase the self-efficacy (Bandura, 1997).

This study supported existing literature addressing the intricacies of guiding pre-service teachers through the complexities of lesson planning (John, 2006; Rusznyak & Walton, 2011). In this study, mentoring built a metacognitive scaffold, which pre-service teachers believed they could use to inform their future planning. Mentoring was the bridge which addressed Fuller's concerns via Bandura's concept of self-efficacy.

Participation was voluntary, and highly motivated pre-service teachers might be more likely to participate in a study, which requires them to dedicate additional time and effort. Additionally, these findings should be viewed as preliminary as they are limited by the small sample size and specificity (secondary education pre-service teachers) of the participants. Like Drost and Levine (2015), the researchers call for continued examination of the topic.

Implications

Though learning to lesson plan can be a challenge, as an essential function of teaching,

pre-service teachers must be equipped with the knowledge, skills, and dispositions to do so. Results of this research are compelling enough to consider implementation of lesson plan mentoring into early coursework in teacher preparation programs. The positive outcomes of this intervention indicate the value of mentoring early-program pre-service teachers as they learn how to plan. However, can this experience be replicated without the resources required for intense mentoring?

One way to incorporate mentor-like support might be through the flipped classroom approach. The mentor role in this approach could be assumed by a course instructor. Jones et al. (2011) suggested including intentional lesson planning discussions into the early traditional course requirements. Lesson plan components could be introduced via videos pre-service teachers watch prior to class, thus allowing instructors to spend course time modeling and discussing metacognitive processes involved in lesson planning. Another way to incorporate mentoring might be through adapting the mentor conference used in this study. Course schedules could be adjusted to allow time for instructor-led lesson plan conferences. The conference structure could also be adjusted to include small-group sessions comprised of pre-service teachers with similar needs.

Mentoring during the early stages of teacher preparation programs may be one way to mitigate concerns and increase the self-efficacy of pre-service teachers in regard to lesson planning. As such, teacher preparation programs should consider mentoring as a means to better prepare pre-service teachers for the expectations of their future classrooms.

References

- Ambrosetti, A. (2009). Mentoring and learning to teach: What do pre-service teachers expect to learn from their mentor teachers? *International Journal of Learning*, 17(9), 117-132.
- Ambrosetti, A., Knight, B.A., & Dekkers, J. (2014). Maximizing the potential of mentoring: A framework for pre-service teacher education. *Mentoring & Tutoring: Partnership in Learning*, 22(3), 224-239. doi:10.1080/13611267.2014.926662
- Ashton, P. (1984). Teacher efficacy: A motivational paradigm for effective teacher education. *Journal of Teacher Education*, 35(5), 28-32. doi:10.1177/002248718403500507
- Schaffer, C. (2003). *Changes in student teachers' perceptions of stress during the student teaching semester*, (Doctoral Dissertation), University of Nebraska at Omaha, Omaha, NE.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191-215. doi:10.1037/0033-295x.84.2.191
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W.H. Freeman and Company.
- Bennis, W. (2003). *On becoming a leader*. New York: Basic Books.
- Cho, M. (2017). Pre-service L2 teacher trainees' reflection: What do they focus on? *English Teaching*, 72(1), 105-129.
- Cherubini, L. (2009). Reconciling the tensions of new teachers' socialization into school culture: A review of the research. *Issues in Educational Research*, 19(2), 83-99.
- Constas, M. A. (1992). Qualitative analysis as a public event: The documentation of category development procedures. *American Educational Research Journal*, 29(2), 253-266.
- Council of Chief State School Officers (CCSSO). (2013, April). *Interstate teacher assessment and support consortium InTASC model core teaching standards and learning progressions*

- for teachers 1.0: A resource for ongoing teacher development*. Washington, DC: Author.
- Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage.
- Creswell, J. W. (2005). *Educational research: Planning, conducting and evaluating quantitative and qualitative research*. Upper Saddle River, NJ: Merrill.
- Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., & Hanson, W. E. (2003). Advanced mixed methods research designs. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 209–240). Thousand Oaks, CA: Sage.
- Dembo, M.H., & Gibson, S. (1985). Teachers' sense of efficacy: An important factor in school improvement. *The Elementary School Journal*, 86(2), 173-184. doi:10.1086/461441
- Denzine, G.M., Cooney, J.B., & McKenzie, R. (2005). Confirmatory factor analysis of the teacher efficacy scale for prospective teachers. *The British Journal of Educational Psychology*, 75, 689-708. doi:10.1348/000709905x37253
- De Winter, J. C. (2013). Using the student's t-test with extremely small sample sizes. *Practical Assessment, Research and Evaluation*, 18(10).
- Drost, B. D., & Levine, A. A. (2015). An analysis of strategies for teaching standards-based lesson plan alignment to preservice teachers. *Journal of Education*, 195(2), 37-47.
- Eun Kyung, K. (2012). What is your objective?: Preservice teachers' views and practice of instructional planning. *International Journal of Learning*, 18(7), 89-100.
- Feiman-Nemser, S., & Rosaen, C. (1997a). *Guiding teacher learning: Insider studies of classroom work with prospective & practicing teachers*. Washington, DC: AACTE Publications.
- Feiman-Nemser, S., & Rosaen, C. (1997b). *Guiding teacher learning: A fresh look at a familiar*

- practice. In S. Feiman-Nemser & C. Rosaen (Eds.), *Guiding teacher learning: Insider studies of classroom work with prospective & practicing teachers* (pp. 8-36). Washington, DC: AACTE Publications.
- Fielder, C. (2014). Are detailed objectives really necessary in lesson planning? *Teacher Trainer: A Practical Journal for Those Who Train, Mentor and Educate TESOL Teachers*, 28(2), 18-20.
- Fuller, F. (1969). Concerns of teachers: A developmental conceptualization. *American Education Research Journal*, 6, 207-226. doi:10.3102/00028312006002207
- Greene, J. C., & Caracelli, V. J. (1997). Defining and describing the paradigm issue in mixed-method evaluation. In J.C. Greene & V.J. Caracelli (Eds.), *Advances in mixed-methods evaluation: The challenges and benefits of integrating diverse paradigms*. San Francisco: Jossey-Bass. doi:10.1002/ev.1068
- Hobson, A. J., Ashby, P., Malderez, A., & Tomlinson, P. D. (2009). Mentoring beginning teachers: What we know and what we don't. *Teaching and Teacher Education*, 25(1), 207-216.
- Hobson, L.D., Harris, D., Buckner-Manley, K., & Smith, P. (2012). The importance of mentoring novice and pre-service teachers: Findings from a HBCU student teaching program. *Educational Foundations*, Summer-Fall, 67-80.
- Hoy, A.W., & Spero, R.B. (2005). Changes in teacher efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education* 21, 343-356. doi:10.1016/j.tate.2005.01.007
- Hudson, P., & Nguyen, T.M.H. (2008, November 30 - December 4). *What do preservice EFL teachers expect from their mentors?* Paper presented at the Annual Conference of the

Australian Association of Research in Education. Brisbane, Australia.

Hudson, P., Usak, U., & Savran-Gencer, A. (2009). Employing the five-factor mentoring instrument: Analysing mentoring practices for teaching primary science. *European Journal of Teacher Education*, 32(1), 63-74. doi:10.1080/02619760802509115

Ingersoll, R., Merrill, L., & May, H. (2012). Retaining teachers: How preparation matters. *Educational Leadership*, 69(8), 30-34.

John, P. D. (2006). Lesson planning and the student teacher: Re-thinking the dominant model. *Journal of Curriculum Studies*, 38(4), 483-498. doi:10.1080/00220270500363620

Johnson, A. P. (2000). It's time for Madeline Hunter to go: A new look at lesson plan design. *Action in Teacher Education*, 22(1), 72-78. doi: 10.1080/01626620.2000.10462994

Jones, K. A., Jones, J., & Vermette, P. J. (2011). Six common lesson planning pitfalls - recommendations for novice educators. *Education*, 131(4), 845-864.

Kelehear, Z. (2003). Mentoring the organization: Helping principals bring schools to higher levels of effectiveness. *NASSP Bulletin*, 87(637), 35-47. doi:10.1177/019263650308763704

Kouzes, J. M., & Posner, B. Z. (2002). *Leadership the challenge*. San Francisco: Jossey-Bass.

Lin, H., & Gorrell, J. (2000). Exploratory analysis of pre-service teacher efficacy in Taiwan. *Teaching and Teacher Education*, 17(5), 623-635. doi:10.1016/s0742-051x(01)00018-x

Morton, T., & Gray, J. (2010). Personal practical knowledge and identity in lesson planning conferences on a pre-service TESOL course. *Language Teaching Research*, 14(3), 297-317. doi:10.1177/1362168810365243

Northouse, P. G. (2004). *Leadership theory and practice*. Thousand Oaks, CA: Sage.

Orland-Barak, L. (2014). Mediation in mentoring: A synthesis of studies in *Teaching and Teacher Education*, *Teaching and Teacher Education*, 44, 180-188.

doi:10.1016/j.tate.2014.07.011

Orland-Barak, L., & Hasin, R. (2010). Exemplary mentors' perspectives towards mentoring across mentoring contexts: Lessons from collective case studies. *Teaching and Teacher Education*, 26(3), 427-437. doi:10.1016/j.tate.2009.05.009

Peralta, L. L., & Burns, K. (2012). First off the blocks: professional experience and learning for first-year preservice physical and health education teachers. *Asia-Pacific Journal of Teacher Education*, 40(2), 127-141. doi:10.1080/1359866X.2012.669827

Purdum-Cassidy, B. B., Nesmith, S., Meyer, R., & Cooper, S. (2015). What are they asking? An analysis of the questions planned by prospective teachers when integrating literature in mathematics. *Journal of Mathematics Teacher Education*, 18(1), 79-99. doi:10.1007/s10857-014-9274-7

Rusznayak, L., & Walton, E. (2011). Lesson planning guidelines for student teachers: A scaffold for the development of pedagogical content knowledge. *Education as Change*, 15(2), 271-285. doi:10.1080/16823206.2011.619141

Ruys, I., Keer, H. V., & Aelterman, A. (2012). Examining pre-service teacher competence in lesson planning pertaining to collaborative learning. *Journal of Curriculum Studies*, 44(3), 349-379. doi:10.1080/00220272.2012.675355

Schneider, R. (2007). Mentoring new mentors: Learning to mentor preservice science teachers. *Journal of Science Teacher Education*, 19(2), 113-116. doi:10.1007/s10972-007-9088-x

Schwille, S. A. (1997). Louise and me: An analysis of field instructor's practice. In S. Feiman-Nemser & C. Rosaen (Eds.), *Guiding teacher learning: Insider studies of classroom work with prospective & practicing teachers* (pp. 53-71). Washington, DC: AACTE Publications.

Smith, E. R. (2005). Learning to talk like a teacher: Participation and negotiation in co-planning

- discourse. *Communication Education*, 54(1), 52-71. doi:10.1080/03634520500076778
- Stake, R. E. (2005). Qualitative case studies. In N. K. Denzin & Y. S. Lincoln (eds.), *The Sage handbook of qualitative research* (3rd ed., pp. 443-466). Thousand Oaks, CA: Sage.
- Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Thousand Oaks, CA: Sage.
- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Applied Social Research Methods Series, vol. 46, Thousand Oaks, CA: Sage.
- Tummons, J. (2010). The assessment of lesson plans in teacher education: a case study in assessment validity and reliability. *Assessment and Evaluation in Higher Education*, 35(7), 847-857.
- Wasonga, C. O., Wanzare, Z.O., & Dawo, J.I. (2015). Mentoring beginning teachers: Bridging the gap between pre-service training and in-practice realities. *Journal of International Education and Leadership*, 5(2), 1-11.
- Wee, S., Weber, E. K., & Park, S. (2014). Early childhood practicum students' professional growth in the USA: Areas of confidence and concern. *International Journal of Early Years Education*, 22(4), 409-422. doi:10.1080/09669760.2014.911644

Difficulty and Distance in Educational Encounters with Historical Violence

Grant Scribner

Department of Teaching, Learning, and Teacher Education
University of Nebraska-Lincoln

Abstract

This article reviews recent literature that addresses historical violence, difficult history, and the production of historical distance in teaching and learning about past violence. The author argues that based on the literature, the processes by which certain violent histories become “difficult” while others are aestheticized deserve greater attention. As violent histories become more or less difficult, the production of nuanced, contextually contingent historical distances may have serious implications for teachers’ pedagogical decisions as well as students’ reactions and understanding. The author argues further that historical violence not considered difficult or traumatic in a given moment and context deserves greater attention from scholars and researchers.

Keywords: violence, social studies education, history education, assessment, cognition

doi: 10.32873/unl.dc.ne003

The presence of violence in history curriculum is peculiar. Educators face a pedagogical dilemma brought on by the sheer volume and ubiquity of violence throughout human history; it becomes simultaneously inescapable and easily avoided. In elementary and secondary schools--where brief, easily digestible narratives are prized--violence is generally deployed in one of two ways: as a catalytic, explanatory device for the movement of history, or as a whetstone for students to sharpen their capacities for moral judgment. In the first, violence is approached with analytical distance; it demonstrates cause-effect relationships, explains historical processes like the movement of peoples and technological development, bolsters defenses of the state, or pokes holes in the state's logic and tools. In the second, teachers use instances of violence to collapse historical distance through affective connection as they appeal to students' emotions or use historical violence as a filter for refining moral judgment. Missing from both approaches is a careful accounting for the relationship between teachers, violence in curriculum, and the process of creating and representing historical distances with students.

Whichever general approach is taken to accounting for violence, the terms of teachers' and students' encounters with it are by no means straightforward. The process of mediating violent pasts in the creation and enactment of curriculum present educators and students with both productive opportunities and potentially dire risks: to confront or avoid, to valorize or question, to understand violence as "difficult" and traumatic or to blithely subsume it within less troubling narratives. Decisions about whether and how to teach about violence animate much of the literature reviewed here, though it is necessary to understand from the outset that the complexities of these issues have led to a wide range of approaches from scholars with different concerns, vocabularies, and research agendas. The articles in this review have been selected, in part, because of their differences, which serve as a means of triangulating the various ways

historical violence is mediated in curriculum enactment. The problem of violence in history is layered: it begins as an uneven process of historical inscription and distance in the creation of the archive and curriculum, then becomes an issue of intellectual and emotional understanding on the part of the teacher, and finally grows into a set of pedagogical dilemmas as the teacher attempts to engage violent histories with students.

The articles reviewed use a number of similar terms, though often with slightly different meanings and analytical ends in mind: “historical trauma,” “violent pasts,” “difficult history,” and “difficult knowledge.” I use the terms “historical violence” and “violent histories” as a means of maintaining focus on the presence of violence in the remembered past, regardless of its political or relational resonance. Many factors--sociocultural, historical, temporal--influence the degree to which a given event is understood to be difficult or traumatic, but my analytical interest falls in the space between violence understood to be “difficult” and violence treated as banal; both have serious implications for teachers and students as they create and interact with curriculum. I explore the relationship between historical violence and curriculum first by setting out “what makes difficult history difficult” (Gross & Terra, 2018) and then weighing the contributions of scholarship on teaching about violent histories in three contexts: the relationship between historical thinking and violence (with particular attention to issues of recency, proximity, and historical distance), the relationship between teachers and violence in curriculum, and finally, pedagogical concerns and approaches used with students.

Problems of Difficulty and Aestheticization

In David Lowenthal’s (2015) return to his seminal work on conceptions of history, *The Past is a Foreign Country: Revisited*, he argues that thinking about the past presents three potential “evils,” two of which have conditioned how people understand and make use of past

suffering: “the concomitant griefs that the grievous past saddles on the present...and the menace of its continuing potency” (p. 129). The degree to which these concerns become salient is historically contingent--it matters *when* one considers the past as well as *which* past is under consideration--an insight that has profound influence on the lines drawn by Gross and Terra (2018) around what makes difficult history difficult in a given context or moment. Their framework situates violence as one of five criteria necessary to qualify as difficult history, the other four being: centrality to the history of a nation; refuting “broadly accepted versions of the past or stated national values”; relevance to contemporary problems; and creating “disequilibria that challenge existing historical understandings” (p. 54). Gross and Terra argue that a particular type of violence, “usually collective or state sanctioned,” often qualifies as difficult because “violence approved by the state or enacted by groups of supposedly upstanding citizens cannot be easily dismissed as aberrations or exceptions” (p. 54). Similarly, Zembylas and Bekerman (2008) argue that certain memories and forms of social remembrance become “dangerous” only when disruptive to the status quo, “the hegemonic culture of strengthening and perpetuating existing group-based identities” (p. 125).

Simon et al. (2000) offer a slightly different understanding of social remembrance pedagogies as either “strategic practice” or “difficult returns,” both of which they find wanting. Strategic practice refers to “efforts to mobilize attachments and knowledge that serve specific social and political interests within particular spatiotemporal frameworks” (p. 3). The authors note that such efforts have been employed both in service of hegemonic nationalism and that “the remembrance of mass violence has sustained the demand for institutionalized practices of justice” based on an assumption that remembrance is ultimately oriented toward improving future conditions. The authors further argue that “on these terms, no matter how horrible its

stories, a strategic remembrance of mass systemic violence is consolatory...[and] dependent upon a moralizing pedagogy” (pp. 3-4). Pedagogies of difficult returns? rest on a premise of “learning to live with a disquieting remembrance” that makes “the memorial impulse to turn and return traumatic history [into] an *assignment*, not simply a matter of choice” (p. 4). Simon et al. (2000) take issue with pedagogies of strategic practice because they privilege continuity by collapsing historical distance to reshape the violent past for contemporary political or social use. On the other hand, pedagogies of difficult return make unhelpful claims to continuity through “practices of identification that threaten to collapse differences across space/time and through performances of surrogacy” that heighten the risk of paralyzing the living in their traumatic remembrances (p. 5). The authors argue that both approaches ultimately fail to acknowledge adequately the “politics of relationality” between people and historical violence. Simon et al. (2000) offer a remedy by positing a third path: attending to historical violence through a process of “critical learning” that “enacts the possibilities of hope through a required meeting with traumatic traces of the past” (p. 5-7), a productive way to incorporate the strengths of the other two approaches without creating false continuity with historical violence.

The formulations of Simon et al. (2000), Zembylas and Bekerman (2008), and Gross and Terra (2018) make sense for filtering acts of historical violence through lenses of difficulty and trauma, but filtering historical violence on the basis of contemporary contextual resonance or state sanction implies that many—perhaps most—acts of historical violence that appear in curricula year after year may not qualify as “difficult” at a given moment. As Gross and Terra (2018) point out, If the difficulty of history is socially constructed, both outside the classroom and through the interactions of teachers and students with the curriculum, historical violence that is not deemed difficult may represent the majority of history curriculum in some contexts

(particularly in teaching the history of the United States). With that in mind, it is troubling that the great weight of violence throughout history is understood to be (and taught as) untroubling. What, then, are the pedagogical implications of teaching about acts of violence that are not considered difficult or traumatic? Gross and Terra (2018) point out that “educators are sometimes reluctant to tackle...difficult histories in the classroom – and when they do, their instruction may be inadequate” (p. 52). It is unclear whether teachers avoid certain topics because they recognize their own practices as inadequate or simply because of the social discomfort of difficult histories. In either case, the questions remain about teachers’ pedagogical decisions concerning the acts of violence they presumably present to their students that are not understood to be difficult. Does a teacher’s discomfort with (or avoidance of) difficult historical violence imply greater sensitivity when she approaches less troubling moments of violence in curriculum? Or, does avoidance imply that a teacher might treat other acts of historical violence as completely unproblematic? A teacher’s disposition toward the difficulty of certain events or histories would have profound effects on her pedagogical decisions and, in turn, students’ reactions and learning. To limit our attention to “difficult” moments may obscure a broader view of the dispositions teachers adopt toward the majority of historical violence and the effects their dispositions may have on students’ reactions and understanding.

Raudsepp and Zadora’s (2019) analysis of teachers’ sensitivity toward violence in World War II provides some evidence that it is necessary to give greater research attention to historical violence that is not considered difficult, or to the processes by which difficult historical violence becomes less so. Based on the open responses of 719 teachers from across Europe to an online questionnaire about their perceptions of the violence and atrocities of World War II, the authors argue:

“The reasons for sensitivity [among teachers] were found to lie in cognitive and emotional barriers to treating the atrocities, violence and discrepancies between different perspectives on WW2 of social memory groups. A relatively new dimension, confirmed by the research, is the problem of the aestheticisation of the violent past. This can lead to banalisation and even legitimisation of the violence and aspects of the violence that pupils perceive to be fascinating” (p. 87).

While in many respects the findings fit neatly into Gross and Terra’s framework of difficulty as conditioned by the nation state and perceived contemporary relevance, the problem of historical violence’s aestheticization necessitates a closer examination of the processes and relationships that create the intellectual or emotional space (outside or beyond the state) for teachers and students to treat violence as either banal or legitimate. Helmsing (2014) provides useful theoretical grounding for considerations of affective connection in social studies education. He uses critical theory and feminist post-structural theory to argue that two affects present in curriculum materials and enactment—pride and shame—contribute to production of subjects in social studies education. One of Helmsing’s examples, the textual and artistic production of affect in curricular materials through triumphant depictions of U.S. involvement in World War II, shows one common intersection of between teachers, students, historical violence, curriculum, and the state. While a textbook privileges a certain type of affective connection to violence in that instance, the dispositions of the teacher and students—the particular understandings of historical distance they each bring to the textual encounter—must also be investigated and carefully considered.

Different Measures of Historical Distance

In order to understand how teachers and students encounter historical violence in curriculum it is necessary to consider the nature of historical thinking about violence—in particular, the dynamic relationships between temporal distance, historical distance, and affective connection to the past. Phillips (2013) argues that temporal distance is “just the beginning, since

historical understanding is inconceivable outside of the affective and ideological structures that make representation possible. Consequently, an idea so fundamental to the historical vocabulary might need to be rethought in more open and imaginative terms” (p. xi.). As a means of mediating these problems of distance, Phillips proposes the application of a heuristic framework that accounts for four distances: formal, summoning, affective, and conceptual. Phillips’ heuristic attends simultaneously to the forms of historical representation, their affective claims, and the contemporary means of understanding upon which “history’s intelligibility depends” (p. 14). Klein (2017) argues that the complexities of interaction among these forces leads to “longer and shorter distancing in representations of the past,” illustrating a need for an analytical framework “imagined as a continuum, where many intermediate positions and complicated combinations are possible, creating unique varieties of distance” (p. 184).

Raudsepp and Zadora’s (2019) study provides a broad base of data that supports the arguments of Phillips and Klein. By framing the aestheticization of violence as problematic, Raudsepp and Zadora (2019) underscore the degree to which both historical and temporal distance influence students’ affective connections when encountering the past; for the authors, the weakening affective connection over time is troubling, which implies that eliciting some sort of emotional reaction or affective connection from students is (or should be) a pedagogical goal for teachers. They describe the degree of affective connection between students and the atrocities of the Holocaust as “the regulation of emotions,” which teachers relate to “the goal of moral education” (p. 103). The authors argue that students responding to the violence of the Holocaust with empathy and grief were understood to have taken the necessary lesson, while “indifference or fascination with the atrocities were understood [by teachers] to be indications of moral failure” (p. 103).

These insights have particular explanatory power when applied to the shifting resonances of historical violence because the dynamics of historical distance—largely due to the uneven development of national projects and processes of social memory—had profound effects on the pedagogical decisions of teacher participants and the affective responses of student participants. Raudsepp and Zadora offer two speculative explanations for the banalization of historical violence. The first is socio-political context: one Arab Israeli teacher, for instance, identifies the Holocaust as a “sensitive issue” that needs to be taught, but argues that students’ perceptions of the approved curricular materials as politically biased leads them “to become apathetic and discredit the claims of the material” (p. 101). The second explanation is that cultural change may alter the nature of students’ interactions with the violent past: “The expansion of historical culture to the digital sphere provides young people with novel forms of engagement with historical events, such as, videogames” (p. 88).

The second explanation, while certainly plausible, is not fleshed out, which leads back to a broader assessment of the state of the research field. Laying out the trajectories of research on historical violence, Carretero (2017) nods to the affective and representational problems of historical distance in a call for more unified scholarly discourse:

As it can be easily imagined, it is not the same to teach, for example, the history of Roman Empire than to discuss in the classrooms about a national civil war that happened one or two decades ago. In this respect, most of the present advances on history learning and teaching have to do with how to teach and learn historical contents [sic]. On its part, most of the research on history textbooks has to do with what is included in the textbooks...[In] the area of research history education, and particularly in the field of the role of history education for conflict resolution and reconciliation, we need to establish a more meaningful relation between these two areas of research. (pp. 346-347).

Carretero (2017) does not explicitly address the issue of violence as bound up in the difficulties of historical distance (using “conflict” instead), but those issues certainly influence

his argument for understanding and treating historical conflict in three separate contexts: academic history, school history, and popular history. While the interactions of those three types of historical discourse condition teacher decisions and student reactions to violence, for Carretero the key to understanding the “difficulty” of violence is uncovering the relationship between individuals and the state, an argument echoed by Gross and Terra (2018).

As Carretero notes, textbooks provide an obvious intersection between history and the state; they are sites of mediating historical distance that also present a manageable evidence base for researchers. In one study of historical violence in textbooks, Brown and Brown (2010) use critical race theory and the concept of cultural memory to examine how racial violence in U.S. history is framed in twelve contemporary textbooks. They argue that the texts present violence against African Americans as both organized and often premeditated but also “deinstitutionalized acts undertaken by ‘bad’ men or ‘bad’ people” (p. 44). That insight leads the authors to argue:

“Although accounts of racial violence that historically have been excluded from textbooks are now being included, this inclusion matters little if it is presented in a manner that disavows material implications of racial violence on sustained White privilege and entrenched African American inequities” (p. 31).

If we apply Phillips’s (2013) heuristic as a means of understanding further curricular implications of the texts’ representations of historical violence, Brown and Brown’s argument certainly attends to the affective and conceptual elements of historical distance. The violence in the textbooks studied is portrayed as brutal and unjust, collapsing affective distance for students; simultaneously, violence is assigned to “bad” individuals, widening the conceptual distance between injustice, organized violence, and the American state or American society. The resulting historical distance invites students to feel past violence keenly while directing any negative reactions or judgments toward individuals and away from the state and society. In that sort of

historical representation the risks of affective connection--encounters that become traumatic for students--are outweighed by the potential social benefits of distancing historical violence from the contemporary state.

Pedagogical Choices and Student Understanding

In a similar vein, scholars and teachers have given attention to mediating historical violence through pedagogical decisions attempting to understand student reactions to their encounters with historical violence. Encounters with difficult knowledge have been framed productively as pedagogical and research dilemmas that are influenced by the past experiences of students and their relationships to the teacher/researcher. Psychoanalytic processes of uncovering offer one path toward understanding the dynamics of these interactions (Pitt & Britzman, 2003; Collins, 2013), though as Gross and Terra (2018) note, it may be unrealistic to expect practicing teachers to also be experts in psychoanalysis. Concerns about temporal distance and spatial proximity of violence to students are also central to teachers' pedagogical decisions and their subsequent understandings of student reactions to historical violence. In studies of student encounters with recent historical violence teacher conceptions of temporal and spatial proximity to historical violence increased sensitivity to students' affective distance and influenced pedagogical decisions (Moyo & Gonye, 2015; Lauritzen & Nodeland, 2017). However, as Lauritzen and Nodeland (2017) note, students in one study appear to have internalized a different message than teachers had in mind; students described material support given by the school in the wake of relatively recent community violence, rather than reporting the emotional support teachers had attempted to embed in their own pedagogies.

Conclusion

What should we make of these various conceptions of the relationship between historical distance, violence, and students' encounters with the past? Researchers have framed historical violence using assignments of difficulty based on shifting resonances, perceptions of moral obligation, and the agenda of the state (Gross & Terra, 2018). Researchers have also argued for constructions of historical distance either as a continuum (Klein, 2017) or heuristic (Phillips, 2013) for understanding contemporary relationships to past violence. Affective connections to historical violence may be framed as a worthy pedagogical goal, moral obligation, or dire risk--the appropriate strength or weakness of those connections to be weighed differently in each case (Simon et al., 2000). With those connections in mind, researchers have pointed toward the emotional sensitivities of students and teachers as guides for pedagogical decision-making in ways that privilege recency, proximity, and past personal experiences/encounters (Pitt & Britzman, 2003; Collins, 2013; Moyo & Gonye, 2015; Lauritzen & Nodeland, 2017). However, the risk of pinning constructions of resonant historical violence too closely to temporal recency--the aestheticization and banalization of more distant historical violence (Raudsepp & Zadora, 2019) raises questions about how, or whether, researchers and teachers account for historical violence as it appears to become less salient or resonant. Does such violence simply recede from social memory? Does it become more pliable and easily shaped toward the ends of the state? In short, the dynamic process of seeing and confronting historical violence as difficult history, also a social and curricular process of forgetting--of easing the difficulty present in some historical violence. Such a process, as it plays out in the teaching and learning of history in schools, may deserve more careful attention.

References

- Carretero, M. (2017). The teaching of recent and violent conflicts as challenges for history education. In C. Psaltis, M. Carretero, & S. Čehajić-Clancy (Eds.), *History education and conflict transformation* (pp. 341-377). Palgrave Macmillan.
- Collins, A. (2013). Teaching sensitive topics: transformative pedagogy in a violent society. *Alternation*, 9, 128-149.
- Gross, M. & Terra, L. (2018). What makes difficult history difficult? *Phi Delta Kappan*, 99 (8), 51-56.
- Helmsing, M. (2014). Virtuous subjects: A critical analysis of the affective substance of social studies education. *Theory & Research in Social Education*, 42 (1), 127-140.
- Klein, S. (2017). Preparing to teach a slavery past: History teachers and educators as navigators of historical distance. *Theory & Research in Social Education*, 45(1), 75-109. DOI: 10.1080/00933104.2016.1213677
- Lauritzen, S. M., & Nodeland, T. S. (2017). What happened and why? Considering the role of truth and memory in peace education curricula. *Journal of Curriculum Studies*, 49(4), 437–455. <https://doi.org/10.1080/00220272.2016.1278041>
- Lowenthal, D. (2015). *The past is a foreign country: Revisited*. Cambridge University Press.
- Moyo, N., & Gonye, J. (2015). Representations of ‘difficult knowledge’ in a post-colonial curriculum: Re-imagining Yvonne Vera’s *The Stone Virgins* as a ‘pedagogy of expiation’ in the Zimbabwean secondary school. *Pedagogy, Culture & Society*, 23(3), 455-475.
- Phillips, M.S. (2013). *On historical distance*. Yale University Press.
- Pitt, A. & Britzman, D. (2003). Speculations on qualities of difficult knowledge in teaching and learning: an experiment in psychoanalytic research. *International Journal of Qualitative Studies in Education*, 16(6), 755-776.
- Raudsepp, M., & Zadora, A. (2019). The sensitive scars of the second world war in teaching European history. *Pedagogy, Culture & Society*, 27(1), 87-110.
- Simon, R. I., Rosenberg, S., & Eppert, C. (Eds.), (2000). *Between hope and despair: Pedagogy and the remembrance of historical trauma*. Rowman & Littlefield Publishers, Inc.
- Zembylas, M. & Bekerman, Z. (2008). Education and the dangerous memories of historical trauma: Narratives of pain, narratives of hope. *Curriculum Inquiry*, 38(2), 125–54. <https://doi.org/10.1111/j.1467-873X.2007.00403.x>.

Culturally Relevant Science Teaching: A Literature Review

Uma Ganesan

Department of Teaching, Learning, and Teacher Education

University of Nebraska-Lincoln

Abstract

This educational research literature review paper aims to discuss the rationale, review eight empirical research studies, and identify knowledge gaps in culturally relevant pedagogy in science education. This paper focuses on synthesis, review, and comparison of the findings of the empirical studies, and categorizes them into thematic heads such as similarities and differences between studies under the broad categories of professional development (PD) programs and case studies. Following these reviews, the author summarizes her reflections and thoughts about the literature to understand the big picture of culturally relevant pedagogy in science education. The basis of this literature review are various philosophical foundations that undergird the research in this field, looking at the theoretical frameworks and standards in science education, such as Next Generation Science Standards. This paper hopes to identify knowledge gaps for future research and help educators address serious and pressing concerns regarding culturally relevant science teaching in an increasingly diverse world.

Keywords: Culturally relevant pedagogy, science education, professional development programs, equitable access to education, student science achievement

doi: 10.32873/unl.dc.ne004

Culture is the focal point of learning. Individual cultural experiences play an important role not only in receiving and communicating information, but also in molding the thinking process of individuals and groups of people. As a science teacher, I have always had a strong urge to adopt a pedagogy that acknowledges, responds to, embraces, and celebrates all cultures. I have hoped that this pedagogy should offer full, equitable access to education for students from all cultures. Culturally responsive teaching is a pedagogy that recognizes the importance of including students' cultural references in all aspects of learning (Ladson-Billings, 1994).

Because of my passion for science education in the multicultural perspective and the growing need for educators to embrace culturally relevant teaching in an increasingly diverse world, this topic “culturally responsive pedagogy and culturally relevant teaching of science,” is an extremely important one in the field of education at the present time. This educational research literature review paper aims to discuss the rationale, review eight empirical research studies, and identify knowledge gaps in culturally relevant pedagogy in science education. I have synthesized, reviewed, and compared the findings of empirical studies and categorized them into thematic heads such as similarities and differences between studies under the broad categories of professional development (PD) programs and case studies. Following these reviews, I have summarized the overall findings of the literature, and I have also tried to understand the big picture of culturally relevant pedagogy in science education. I have based my literature review on various foundational studies in this field, looking at the theoretical frameworks and standards in science education, such as Next Generation Science Standards. I plan to incorporate what I have learned from these studies in my future research which will help me to understand deeply about culturally relevant pedagogy in science education. Finally, I plan to conclude this paper by identifying knowledge gaps, listing lingering questions, and understanding how these questions

will help structure future research and streamline my own research on culturally relevant science teaching. In the two subsections of the introduction, I plan to propose my rationale for my research topic highlighting the philosophical foundations that undergird my research, as well as outline the importance of culturally relevant pedagogy in science education.

Rationale for the Research Topic and Philosophical Foundations

As a science teacher, I have always believed in inquiry-based and hands-on science experiments to increase student engagement and achievement. Additionally, I am equally committed to incorporate culturally relevant/responsive pedagogical and teaching methods for improving the learning outcomes of the increasingly diverse student population, especially the underrepresented student population, such as people of color, girls, LGBTQ+ community, immigrants, ethnic minorities, and indigenous people.

In 1994, Gloria Ladson-Billings, a researcher in the field of education, first described culturally relevant teaching. Culturally relevant pedagogy “empowers students intellectually, socially, emotionally, and politically by using cultural referents to impart knowledge, skills, and attitudes” (Ladson-Billings, 1994, p. 18). Ladson-Billings (1995a) explains that culturally relevant teaching involves three dimensions: academic success, development of critical consciousness, and maintenance of cultural competence. For accomplishing student outcomes, every teacher should be knowledgeable about these dimensions; however, it may not be the case for some teachers. Culturally relevant teaching and multicultural education requires a strong knowledge base in the teaching methods with respect to cultural diversity (Gay, 2002). In her book, Sonia Nieto also talks about the need for understanding the socio-political contexts of teaching in diverse classrooms (Nieto, 1996). Ladson-Billings (1995b) further emphasized that teachers should not only motivate their students for academic achievement and cultural

competence, but they must also aid in students' recognition, understanding, and critiquing of ongoing inequities in the society.

I am also inspired by another foundational work by Mary Atwater and Joseph Riley about multicultural science education. Multicultural science education is a field of inquiry with constructs, methodologies, and processes aimed at providing equitable opportunities for all students to learn quality science (Atwater & Riley, 1993). One more noteworthy present-day indigenous researcher is Pauline W. U. Chinn, who has contributed a lot for the development and implementation of culturally relevant, standards-based science curricula for Native Hawaiian students. In one of her studies, pre- and in-service teachers lived with the Native Hawaiian teachers and worked together with them for a period of one year. They culturally immersed with the local population and used their knowledge to develop unique science curricula (Chinn, 2006). Culturally relevant pedagogy also provides a formalized tool for reconciling the standards to include students' native cultures (Emdin, 2011). After the groundbreaking work of the pioneers, there were many other researchers, such as Carla Johnson, Gloria Boutte, George Lee Johnson, and Charlease Kelly Jackson who have worked on culturally relevant practices for teaching science. In the next subsection, I will talk about the importance of culturally relevant pedagogy in science education.

Importance of Culturally Relevant Pedagogy in Science Education

In today's world, there are many challenges faced by schools and colleges in the US because the student population is becoming more and more ethnically and racially diverse, and this is compounded by the fact that there is also a decline in the diversity of the teacher population. In this scenario, culturally relevant teaching has become one of the most important educational considerations in the US and around the world, with ethical implications. The

national science education standards (National Science Teachers Association, 2003) also defines scientific inquiry as “the diverse ways in which scientists study the natural world and propose explanations based on the evidence derived from their work” (National Research Council, 2000, p. 23). In addition, in the US, Next Generation Science Standards were created “by the States for the States” (Next Generation Science Standards, 2013, 2019). Within the NGSS, there are three important dimensions to learning science, namely disciplinary core ideas, science and engineering practices, and crosscutting concepts to help students develop a coherent and scientifically based view of the world around them.

A goal for developing the NGSS was to create a set of research-based, up-to-date K–12 science standards. These standards give local educators the flexibility to design classroom learning experiences that stimulate students’ interests in science and prepares them for college, careers, and citizenship. The NGSS were developed by states to improve science education for *all* students (Next Generation Science Standards, 2013, 2019).

So, to incorporate culturally relevant pedagogy, the traditional forms of curriculum, instruction, and assessment methods need to be revised or altered (Boutte, Kelly-Jackson, & Johnson, 2010). In addition, the students belonging to the underrepresented cultural groups can achieve a higher level of learning. Calling them “scientists” irrespective of their age increases their self-confidence, provides activities and fosters their science identities, and makes them believe that they can achieve anything. Incorporating examples, data, photos, and information of scientists or researchers from different cultures will help reinforce and institutionalize a strong multicultural science education program in schools, colleges, and universities. This will help all students to connect and engage with the content in science classrooms.

Goal of Literature Review and Research Question

The goal or purpose of my literature review of culturally relevant science teaching is to understand the existing literature (peer-reviewed studies) and present their findings in a logical and organized written report. When I analyzed these articles, my aim was to build my knowledge in culturally relevant science pedagogy, understand the important concepts, know about common research methods and experimental techniques, and learn how to apply those concepts to real-life educational settings. The overarching and guiding research question of my literature review is as follows: In what ways can adopting culturally relevant practices help in teaching science to ethnically, culturally, and linguistically diverse students?

Thematic Categorization

To answer my research question, I had selected eight empirical studies to review the culturally relevant science pedagogical strategies used by researchers. The rationale for selecting these studies was that they all had similar theoretical underpinnings, and they all encompassed similar ideologies. The overall aim or focus of my literature review is to provide a comprehensive body of knowledge about culturally relevant science teaching and help readers understand the importance of adopting culturally relevant practices in teaching content-heavy subjects such as science.

Seven of these studies were carried out in the US, and one in New Zealand. The student population in these studies included many ethnic, cultural, and linguistic groups such as Native American, Maori, Hispanic, Latinx, and African American. All the studies were based on concepts drawn from philosophical foundations of Ladson-Billings' culturally relevant pedagogy, Vygotsky's social constructivism, and Freire's critical pedagogy. One of the studies done by Luft, Bragg, and Peters (1999) is based on the framework laid by another great

educational researcher, Marilyn Cochran-Smith. These frameworks foster richer and deeper understanding of working with diverse student communities. While categorizing the studies into themes, I found many commonalities, similarities, and differences between the studies. Four studies, namely Johnson (2011), Grimberg and Gummer (2013); Tolbert (2015); and Ramirez, McCollough, and Diaz (2016) highlighted the benefits of using culturally relevant science pedagogical strategies during either professional development programs or learning events for preservice teachers. The other four studies, Luft et al. (1999), Patchen and Cox-Petersen (2008), Laughter and Adams (2012), and Morales-Doyle (2017) were all case studies, where the researchers used culturally relevant science pedagogical strategies and reported positive outcomes in the students and teachers who participated in the studies. For easy comparison, summarization of the results, and to understand the significance of the studies, I have categorized the major themes into similarities and differences where I will compare two studies at a time so that reviews of all the eight studies will be covered in this paper.

Similarities in Research on PD Programs

Effective professional development programs enable educators to develop the knowledge and skills they need to address students' learning challenges. In the selected empirical studies, I found that Grimberg and Gummer's (2013) and Johnson's (2011) studies were very similar because both focused on science teachers participating in professional development programs and how these programs affected their science teaching practices for different ethnic or cultural groups. In Grimberg and Gummer's study, which was conducted over three years, the focus of the study was on a professional development (PD) program for science teachers who belonged to "25 K-8 schools near or on the reservations of the Native American Indian tribes in the Cheyenne, Crow, and Flathead Reservations in Montana" (Grimberg & Gummer, 2013, p. 19).

Similarly, Johnson's study (2011) was a longitudinal 3-year study focusing and following two science teachers of low-performing, urban-district middle schools that had a growing Hispanic, ELL student population. They had participated in a PD program utilizing the "transformative professional development (TPD) model" (Johnson, 2011, p. 174). The data collection methods of both studies were also identical, mainly classroom observations, surveys, and interviews. Findings of Grimberg and Gummer indicated that after the teachers completed two years in the program, they changed their practices of teaching and what they believed to be their science-teaching ability. This change had resulted in implementation and practice of equitable instruction which had a positive impact on students' performance. Similarly, Johnson's findings showed that transformational PD programs helped the participant teachers use culturally relevant science pedagogical techniques to transform their teaching practice that resulted in more productive instructional surroundings for their ELL students.

The next two studies based on PD programs that I am going to compare are those of Tolbert (2015) and Ramirez et al. (2016). Even though both studies were conducted in different parts of the world, there were commonalities. The study conducted by Ramirez et al. in a southwestern city of the US with a large, fast-growing Latino population, describes the implementation of culturally relevant science and math content program by "preservice teachers (PSTs) at Family Math/Science Learning Events (FM/SLEs)" (Ramirez et al., 2016, p. 43). This event informed the PSTs about the importance of connecting with Latino families' language and culture both in- and after-school environments. Data collected were questionnaires filled out by the PSTs before the event, interviews of the PSTs after the event, recorded interactions between the PSTs and parents during the event, and interviews with non-English speaking Latino parents after the event. Similarly, Tolbert's study shares the results of an impactful professional

development (PD) program called Te Kotahitanga (TK) conducted in New Zealand. The participants of the study included four science teachers and four mentors (facilitators) at four different schools for a period of one year focusing on Year 9 and Year 10 science classrooms. TK program was an educational research development project that had remarkably influenced how secondary schools had successfully retained Maori students while also increasing their participation and achievement. The researcher investigated how TK mentors engaged the science teachers (novice and experienced) in “reflective conversations around culturally sustaining equitable science instruction for indigenous students” (Tolbert, 2015, p. 1325). The data collected were videotaped classroom observations and recorded mentoring conversations between the science teachers and the TK mentors, which were transcribed and coded. Semi-structured individual interviews of the science teachers and the TK mentors and reflections from both the groups were also done and transcribed. Data analysis was done to look for “culturally sustaining pedagogy in science” by analyzing video recordings and coding the transcripts of the interviews and mentoring conversation (Tolbert, 2015, p. 1339). Results of both studies confirm that PD programs do influence the student and teacher outcomes. For example, in the study of Ramirez et al., results strongly indicated that preservice science/math teachers’ perceptions of Latino parents can be changed by participating in these types of events. Similarly, in Tolbert’s study, the findings indicated that mentoring conversations can be powerful tools in helping teachers become better educators of minoritized students in science.

Differences in Research on PD Programs

Research studies were different in a number of ways too. Grimberg and Gummer’s (2013) study is a quantitative study, where the students in focus were Native Americans. Whereas, Johnson’s (2011) study, which is a qualitative one, focused on Hispanic students. In

Grimberg and Gummer's study, the methods of quantitative analysis used were ANOVA and multiple regression. The researchers found that when the teachers employed and believed in strategies focused on equity, the students were motivated to connect issues of real-life science with hands-on experiments; this explained the variance (36.7%) in the students' science test scores between treatment classrooms and control classrooms. In Johnson's study, the qualitative method of analysis used was coding. All the data collected were coded using broad themes such as teachers' beliefs in students' success, community-development activities in schools, developing critical scientific thinking in students using CRP, scaffolding of instruction for students, and using various techniques to assess learning and understanding of students.

In the other two research studies, Ramirez et al. (2016) and Tolbert (2015), the main difference is that they were conducted in completely different settings, the former in an urban southwestern city of the US and the latter in New Zealand. Ramirez et al. focused on Latino students, and in this research, preservice teachers learned that their perceptions about the Latino parents changed by attending Family Math/Science Learning Events. Whereas, Tolbert's study focused on Maori students, and this research confirmed that mentoring science teachers helped them become better teachers of indigenous students in a culturally sustaining environment.

Similarities in Case Study Research

I found that the two case studies of Luft et al. (1999) and Morales-Doyle (2017) were quite similar in location and method because both were conducted in urban western cities of the US, and both were qualitative studies. In the study of Luft et al., the predominant student body was Hispanic American, with a few African Americans and Native Americans. Similarly, in Morales-Doyle's study, the study subjects were nine students of color (African American and Latinx). Luft et al.'s study examines teaching experiences of a student teacher who is an avid

enthusiast of multicultural science education. She wanted to incorporate culturally relevant, inquiry-based, science instruction inclusive of all her students “consistent with the National Science Education Standards” (National Research Council, 1996)” (Luft et al., 1999, p. 528). This case study furthers the research knowledge about student teachers who are learning to teach science in a different cultural setting other than their own. In Morales-Doyle’s case study too, the researcher attempts to address the racial and class-related inequities in science education, which are long standing issues; this study explored the effects of a justice-centered AP chemistry class on students’ academic success while also addressing the critical environmental and social issues of justice determined by local communities. Both of the studies had the theoretical framework of “culturally relevant pedagogy” by Ladson-Billings. The data collected of both studies were similar, such as in-depth interviews, weekly observations by the participants, by studying student artifacts, discussions with the student teacher after the classroom observations of science teaching, and reflective journal entries of the student teacher or the researcher about their teaching experiences. Even data analysis was identical, which was qualitative data analysis with coding of transcripts of interviews. Results of Luft et al. revealed that the student teacher experienced “an unfamiliarity with her students and their life experiences, a marginalization of herself as she tried to create new lessons for students in science, and a desire for her science instruction to be more relevant to her students” (Luft et al., 1999, p. 527). The student teacher did not receive any support from her colleagues and felt marginalized and constrained. The complexities of learning to teach in culturally different environments are revealed in this research. Quite similarly, the findings of Morales-Doyle’s study revealed that science curriculum organized around an environmental racism issue augmented academic achievement of students greater than what is expected in a typical high school AP chemistry course. The findings also

emphasized how the justice-centered curriculum gave the students opportunities to go beyond academic achievement to becoming transformative intellectuals, who can display complex critical thinking about social justice and scientific issues. It also made them committed towards their own communities and cultures and lent credibility to them.

Differences in Case Study Research

There were two case studies which were different in their approaches, namely Laughter and Adams (2012) and Patchen and Cox-Petersen (2008). Laughter and Adams's study was conducted in an urban Title I school in a southeastern city of the USA. The school had students from "low-income neighborhoods" and 51% of its students were from "low socioeconomic status" (Laughter & Adams, 2012, p. 1118). The school had predominantly white students (76%) and the other students were racially diverse. There was also a growing ELL student population. The study focused on culturally relevant science teaching employed by a student teacher, teaching science in five classrooms of sixth graders (average of 30 students in each class) to incorporate the issues of bias and social justice into her science lessons. Whereas, Patchen and Cox-Petersen's case study was conducted in an elementary school in an urban western city of the US. Two teachers, who teach science for classrooms with 100% Hispanic and African American students from grade 2 through grade 4 were selected for this case study (Patchen & Cox-Petersen, 2008). The main premise of Laughter and Adams's study was culturally relevant pedagogy (CRP) by Ladson-Billings (1995). Whereas, Patchen and Cox-Petersen's study borrows frameworks from "two theories, social constructivism (SC) by Vygotsky and culturally relevant pedagogy (CRP) by Ladson-Billings" (Patchen & Cox-Petersen, 2008, p. 995). Both the studies had different viewpoints with respect to their findings. Laughter and Adams based their study on the three tenets and three goals of CRP for which all teachers should aim for, and the

findings revealed that the student teacher's lesson did substantiate the tenets of culturally relevant science teaching for academic success, socio-political consciousness, and cultural competence of students. However, Patchen and Cox-Petersen used theories of social constructivism and culturally relevant pedagogy because these two theories are identified as the mechanisms for diminishing the "disparities in science education" (Patchen & Cox-Petersen, 2008, p. 994). Their findings indicate that the two teachers successfully modified their science teaching practices to better meet the needs of their ethnically diverse students.

Summarization of Major Findings and Limitations of the Selected Literature

When I attempted to review eight empirical studies, I wanted to accomplish two main things: a) consolidate my understanding of the theoretical frameworks of culturally relevant pedagogy in science education, and b) study these primary research studies in detail to understand them and explore whether these studies can answer some of my questions about multicultural science education and culturally relevant pedagogies. I want to convey my thoughts and opinions in the ensuing paragraphs, after having compared all the studies with each other. I would also like to point out some of the limitations of the studies which could lead to future research.

In Grimberg and Gummer's (2013) study, the findings reiterated the point that when students are taught in a culturally responsive way, they learn better, and their test scores improve. In addition, the teachers' beliefs and culturally relevant teaching practices also improved after this PD program. I learned that the combination of local tribal/cultural practices and science teaching methods utilizing the collective wisdom of the community had resulted in the success of the PD program. The limitation of this study was that it was done for a period of 5 years where demographic changes in the teacher and student population could have affected the result.

Another limitation could be that most of the teachers (79%-89% in both the cohorts) were females, and gender-related differences in teaching beliefs and practices could be a matter of contention in these types of PD programs.

Since my focus of my research is going to be on using CRP in science education, Johnson's (2011) study provided great insights about the foundational principles and components that I must focus on in my own research. I learned that using CRP is quite beneficial to teachers who teach ELL students. In addition, as the Hispanic student population is rising in the United States, we need more research studies to understand the effectiveness of instruction for this group of students. However, some limitations do exist while attempting these kinds of studies. TPD is a "time-intensive program"; this study required over "300 hours of support for teachers" and consistent and committed support from the school district to reform the teachers' practices (Johnson, 2011, p. 171). This kind of support and cooperation from institutions may or may not be possible for all teachers of all school districts. Other limitations are the scope, validity, and reliability of this study where only two teachers were followed. So, extrapolating the results of just two subjects to the whole teaching community might not be appropriate because qualitative studies have transferable, but not generalizable, findings.

From Laughter and Adams's study, I learned a lot because by just tying up a single lesson with issues of racism and social justice, the science teacher was successful in eliciting encouraging responses from students which adhered to the tenets of CRP. So, there is a lot of potential for incorporating CRP into daily science lessons. I learned that CRP used in science education can develop students' insights about the importance of science in their own lives, and it can make students understand that they can make a difference in society with their scientific knowledge. The teachers of STEM education could also support issues of social justice by using

the scientific method and problem-solving techniques. However, an important distinction and a disappointing limitation of this study was that this study focused on a single lesson completed in just three days, instead of a wider incorporation into the sixth-grade science curriculum of the school. So, although this study is temporally and spatially limited, it has reiterated the beneficial effects of CRP in science education.

I learned from Luft et al.'s (1999) study that there were four implications identified for teacher educators from this study: future teachers should become familiar with the culture of the students they are likely to instruct; preservice teacher education programs should provide pragmatic approaches to student teachers to work effectively in multicultural classrooms; student teachers should be paired up with conducive schools and cooperating teachers who allow the student teacher to pursue with multicultural science teaching practices; and the student teacher should have a close "other" or a confidante (such as student's peers and instructors) so that the student teacher can confide in them and consider them as guides and reflective partners. I also found a limitation that the study focuses on only one student teacher, and it is difficult to extrapolate the findings of this study to the vast teaching community, but the results can be considered transferable to a similar context at the individual level. The study generated more questions than answering them.

For my future research, the findings of Morales-Doyle's (2017) research study are significant since my prime area of interest is culturally relevant science pedagogy focusing on social justice and diversity. The research study has implications for educational researchers, preservice teachers, practicing teachers, and teacher educators who want to incorporate "science education as a catalyst for social transformation" (Morales-Doyle, 2017, p. 1034). However, the limitation of this study is that only nine students were studied. Cochran-Smith and Lytle (2009)

also lament that the teacher education research has not been able to aggregate findings across studies (Morales-Doyle, 2017, p. 1056).

From Patchen and Cox-Petersen's (2008) study, I learned that both SC and CRP theories aided the teachers in this study to better understand their students and modify their classroom practices to suit the needs of their students. The implications of this study show that culturally relevant pedagogy increases access to scientific and technological fields for marginalized students. I also aim to focus on similar goals for my future research. However, the limitation was that students' outcomes were not measured. Only the teachers' practices were studied and found to improve.

Being a science teacher and researcher, I was curious to know about the results of Ramirez et al.'s (2016) study because of its deep connection to culturally relevant science and math pedagogy. I have experienced that interactions with parents remove the inherent biases that may be present in teachers, and the interactions improve the relationships in and beyond the classroom. Implications of this study for programs for teacher preparation include, giving the PSTs ample opportunities to participate in FM/SLEs that teach them how to plan and teach, and incorporate culturally relevant math and science activities in class. Another gain from this study is that it encourages the model of acceptance where the PSTs can identify and revisit the perceptions and misconceptions about the parents of their students from a different ethnic group. The limitation of this study is that it was only concentrating on the Latino parents who don't speak English. However, knowledge base about many other ethnic groups is also lacking, so studies encompassing more diverse participant groups should be done to learn about the overall implications of those studies in the education field.

I learned from Tolbert's (2015) study that the topic of culturally relevant teacher mentoring/PD programs is underrepresented, and less literature is available for science education researchers. Because of the growing diversity in the student population, these types of studies help us understand how to retain and ensure success for indigenous students in science classrooms. The study has further implications for teacher educators for becoming more culturally responsible mentors. One limitation that I identified is that this study did not concentrate on other ethnic groups, probably because it was conducted in an area which has a high Maori population. Even the percentages of school dropout rates were not given in the study. More studies are warranted to learn more about these types of PD programs.

Identification of Research Gaps and Questions for Future Research

Many researchers have added on to our knowledge of culturally responsive teaching. The methods formulated by them could effectively be used to improve the student outcomes in a diverse classroom setting. When reviewing the research studies on CRP, I found that there were many knowledge gaps in the selected literature, as there are not many research studies being done in this field. So, these gaps could lead to questions being posed for future researchers. The purpose of my future research is to develop workable protocols for multicultural science education and qualitatively analyzing them for effectiveness. I am planning on research work that could potentially help present/future science teachers apply culturally relevant pedagogical methods in their classrooms. The workable research questions that I came up with when reviewing the selected literature are as follows:

1. Could gender-related differences in teaching beliefs and practices be a matter of contention in science teachers' PD programs? In what ways could we generalize those findings to a wider, non-gender specific science teacher population?

2. How can we conduct large controlled studies to investigate ideas derived from small-scale studies that used culturally responsive science pedagogy? In what ways can we address research limitations such as scope, validity, and reliability of existing studies in which only two or three teachers' experiences are studied?
3. What are the most challenging hurdles to be overcome to implement CRP in science classrooms? When, how, and where does change to following the CRP protocol typically occur?
4. How do different school contexts support culturally relevant science pedagogy? What implications do these broad trends and school- and teacher-level processes have for teachers' practices and students' opportunities to learn?
5. How can preservice teacher education programs provide pragmatic approaches to student teachers to work effectively in multicultural classrooms? What strategies are to be used by policy makers and administrators to ensure student teachers are paired up with conducive schools and cooperating teachers who allow the student teacher to pursue multicultural science teaching practices? What are the hurdles and challenges for such an operation?
6. What new types of assessment strategies are to be used for assessing students in multicultural classrooms?

Conclusion

After summarizing the major findings of the selected literature, identifying the research gaps, and posing future research questions, I am now confident that my goal of literature review has been fulfilled. I now have a solid background knowledge about the theoretical frameworks

available in this educational domain. In the following paragraph, I would like to present the benefits of using CRP in science education.

Benefits of Culturally Relevant Science Teaching

There are many benefits of culturally relevant teaching. First, research points to a marked increase in academic scores when teachers employ culturally relevant instruction methods (Au & Kawakami, 1994). Second, because many minority students (whose communities have either been oppressed or suppressed in the past) must spend their lives trying to adapt to the dominant culture, employing culturally relevant teaching methods will encourage and support their values and identities. These methods make minority students feel safe and motivate them to excel in their chosen field. Scaffolding of learning for these minority students helps them to be more engaged in learning and to advance in their career paths (Butler, Burnett, Renfrew, Renfrew, & Smith, 2017). Additionally, justice-centered science pedagogy, which is also a significant talking point in education, applies culturally relevant pedagogy to counter inequities (Morales-Doyle, 2017). The culturally responsive research in science education classrooms serves two purposes. First, it brings a resolution to the question of how to respond to the situation of indigenous and ethnic minority students' "underachievement" in science education without the need to engage in deep and meaningful ways with the excluded communities. The second purpose is to domesticate the knowledge of others—to make indigenous knowledge "fit" prevailing views of science and to bring it under the control of others. Indigenous communities and ethnic minority communities will assert their right to direct, not just inform, research programs in science and science education and develop their potential as partners of science and education (McKinley & Gan, 2014). As summarized in the findings of the extant literature, culturally relevant science pedagogy will help students develop stronger identities and new perspectives, help them engage

more in learning science, will increase their critical thinking skills, will improve their learning outcomes, and ultimately help them consider a career in the sciences.

Finally, considering the various knowledge gaps and lack of adequate number of research studies in the field of culturally relevant science pedagogy, extensive research studies are needed to arrive at reliable conclusions, and more detailed research studies for a wider teaching population are warranted too. To consolidate knowledge about the underrepresented ethnic communities, more research studies are needed to understand how to best address the needs of the increasingly diverse student body. These generic problems of incorporating culturally relevant pedagogy for any subject/stream of study is very much applicable to science pedagogy too. Keeping all of this in mind, I hope and aim to contribute towards culturally relevant pedagogy in science education in my future research!

References

- Atwater, M. M., & Riley, J. P. (1993). Multicultural science education: Perspectives, definitions, and research agenda. *Science Education*, 77(6), 661–668. doi: 10.1002/sce.3730770609
- Au K.H., & Kawakami, A.J. (1994). Cultural congruence in instruction. In E.R. Hollins, J.E. King, & W.C. Hayman (Eds.), *Teaching diverse populations: Formulating a knowledge base* (pp. 5–23). Albany: State University of New York Press.
- Boutte, G., Kelly-Jackson, C., & Johnson, G. L. (2010). Culturally relevant teaching in science classrooms: Addressing academic achievement, cultural competence, and critical consciousness. *International Journal of Multicultural Education*, 12(2), 1-20. doi: 10.18251/ijme.v12i2.343.
- Butler, V., Burnett, R., Renfrew, K., Renfrew, K., & Smith, K. (2017, January 23). Kicking it up a notch: Becoming a culturally relevant science educator. Retrieved October 9, 2019, from <https://thenode.biologists.com/kicking-notch-becoming-culturally-relevant-science-educator/education/#comments> .
- Chinn, P. W. U. (2006). Preparing science teachers for culturally diverse students: Developing cultural literacy through cultural immersion, cultural translators and communities of practice. *Cultural Studies of Science Education*, 1(2), 367–402. doi: 10.1007/s11422-006-9014-0
- Emdin, C. (2011). Moving beyond the boat without a paddle: Reality pedagogy, black youth, and urban science education. *The Journal of Negro Education*, 80(3), 284-295. doi: 10.2307/41341134.
- Gay, G. (2002). Preparing for culturally responsive teaching. *Journal of Teacher Education*, 53(2), 106–116. doi: 10.1177/0022487102053002003
- Grimberg, B. I., & Gummer, E. (2013). Teaching science from cultural points of intersection. *Journal of Research in Science Teaching* 50(1), 12–32. doi:10.1002/tea.21066.
- Johnson, C. C. (2011). The road to culturally relevant science: Exploring how teachers navigate change in pedagogy. *Journal of Research in Science Teaching* 48(2), 170–198. doi:10.1002/tea.20405.
- Ladson-Billings, G. (1994). *The Dreamkeepers: Successful teachers of African American children*. San Francisco: Jossey-Bass.
- Ladson- Billings, G. (1995a). But that’s just good teaching! The case for culturally relevant pedagogy. *Theory Into Practice*, 34(3), 159–165. doi: 10.1080/00405849509543675
- Ladson-Billings, G. (1995b). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal*, 32(3), 465. doi: 10.2307/1163320
- Laughter, J. C., & Adams, A. D. (2012). Culturally relevant science teaching in middle school. *Urban Education* 47(6), 1106–1134. doi:10.1177/0042085912454443.

- Luft, J. A., Bragg J., & Peters C. (1999). Learning to teach in a diverse setting: A case study of a multicultural science education enthusiast. *Science Education* 83(5), 527-543. doi:10.1002/(SICI)1098-237X(199909)83:5<527::AID-SCE2>3.0.CO;2-T
- McKinley, E., & Gan, M. J. (2014). Culturally responsive science education for indigenous and ethnic minority students. *Handbook of research on science education*, 2, 284-300.
- Morales-Doyle, D. (2017). Justice-centered science pedagogy: A catalyst for academic achievement and social transformation. *Science Education*, 101(6), 1034–1060. doi: 10.1002/sce.21305
- National Research Council (2000). *Inquiry and the national science education standards*. Washington, DC: National Academy Press.
- National Science Teachers Association (2003). Standards for Science Teacher Education. Retrieved October 9, 2019 from: <http://www.nsta.org/pdfs/NCATE-NSTAStandards2003.pdf>
- NGSS Lead States. 2013. *Next Generation Science Standards: For States, By States*. Washington, DC: The National Academies Press.
- Next Generation Science Standards. (2019, November 5). Retrieved November 30, 2019, from <https://www.nextgenscience.org/> .
- Nieto, S. (1996). *Affirming diversity: The sociopolitical context of multicultural education*. New York: Longman.
- Patchen, T. & Cox-Petersen, A. (2008). Constructing cultural relevance in science: A case study of two elementary teachers. *Science Education* 92(6), 994–1014. doi:10.1002/sce.20282.
- Ramirez, O., McCollough, C. A., & Diaz, Z. (2016). Creating a model of acceptance: Preservice teachers interact with non-English-speaking Latino parents using culturally relevant mathematics and science activities at family learning events. *School Science and Mathematics* 116(1), 43–54. doi:10.1111/ssm.12150.
- Tolbert, S. (2015). “Because they want to teach you about their culture”: Analyzing effective mentoring conversations between culturally responsible mentors and secondary science teachers of indigenous students in mainstream schools. *Journal of Research in Science Teaching* 52(10), 1325–1361. doi:10.1002/tea.21240.

Interval Estimation of Proportion of Second-level Variance in Multi-level Modeling

Steven Svoboda

Department of Education Psychology
University of Nebraska-Lincoln

Abstract

Physical, behavioral and psychological research questions often relate to hierarchical data systems. Examples of hierarchical data systems include repeated measures of students nested within classrooms, nested within schools and employees nested within supervisors, nested within organizations. Applied researchers studying hierarchical data structures should have an estimate of the intraclass correlation coefficient (ICC) for every nested level in their analyses because ignoring even relatively small amounts of interdependence is known to inflate Type I error rate in single-level models. Traditionally, researchers rely upon the ICC as a point estimate of the amount of interdependency in their data. Recent methods utilizing an interval estimation of the amount of interdependency based the proportion of second-level variance between groups have been developed that avoid relying solely upon point estimates. The likelihood of committing a Type I error when using the interval estimation of the proportion of second-level variance remains unknown. The current project addressed this deficiency in knowledge utilizing simulated data to assess the accuracy of a 95% confidence interval estimation of the proportion of second-level variance (CI-PSLV). Standard errors tended to decrease as sample size increased, and the CI-PSLV captured the second level ICC in 95% of replications.

doi: 10.32873/unl.dc.ne005

The recognition of hierarchical data structures that account for the dependence of observations and corresponding methods to analyze them have received considerable amounts of attention in the past few decades (Raykov, 2010). Traditionally, the intraclass correlation coefficient (ICC), or the proportion of variance in an outcome variable that is between groups or contained at the higher levels of the nested data structure, is used to determine whether or not accounting for the hierarchical nesting of participants is necessary (Raudenbush & Bryk, 2002). Unfortunately, widely agreed upon guidelines for interpreting the magnitude of the ICC do not exist (Raykov, 2010). Recent methods have been developed to interpret the proportion of variance at higher levels between groups using a confidence interval estimation procedure rather than relying solely upon point estimates such as the ICC because the magnitude of the ICC is difficult to interpret (Raykov et al., 2016). However, there remains a critical need to investigate the accuracy of these methods and to develop guidelines that may be followed when evaluating sample size requirements for the underlying asymptotic maximum likelihood estimation theory to obtain practical relevance when using an interval estimation procedure (Raykov et al., 2016). The purpose of the current study was to establish a 95% confidence interval estimation procedure for the proportion of second-level variance (CI-PSLV) as a valuable tool for applied researchers to consider when deciding whether or not to account for hierarchical data structures in their samples. Specifically, the current study provided a demonstration of the CI-PSLV and explored its accuracy in various sample sizes under various degrees of second level dependence.

Motivating Context

Physical, behavioral and psychological research questions often relate to hierarchical data systems (Mass & Hox, 2004). Examples of hierarchical data systems include repeated measures of students nested within classrooms, nested within schools and employees nested within

supervisors, nested within organizations. Hierarchical data structures exist in nature whether psychologists and behavioral scientists recognize their existence and account for the nesting of their subjects within higher order units in applied research. Obviously, observations may not be independent in these data structures. Failure to account for hierarchical data structures likely violates the assumption that errors are independent of each other and identically distributed. This violation would result in biased standard errors associated with the regression coefficients, which in turn, leads to an increased Type I error rate and erroneous interpretations of statistical tests (Mass & Hox, 2004; Raudenbush & Bryk, 2002). Standard (single-level) models are not appropriate in these hierarchical data systems because individual observations at the lowest level are not independent (Mass & Hox, 2004). Multi-level modeling (MLM), also known as hierarchical linear modeling, techniques avoid having to meet the independence of observations assumption of single-level regression models by accounting for the interdependence of level one observations due to hierarchical nesting structures (Raykov, 2010). However, researchers must adhere to financial budgets and time limits that may render multi-level models infeasible due to the costs associated with collecting additional data from higher nested levels. Because of these limitations, it is imperative that applied researchers consider the proportion of variance in an outcome associated with a given level when deciding how to analyze their hierarchical data.

Intraclass Correlation Coefficient

The intraclass correlation coefficient (ICC) is commonly used to provide researchers with an estimate of the amount of interdependence due to hierarchical nesting structures. The ICC informs researchers' decisions when choosing between MLM techniques and single-level modeling (Mass & Hox, 2005; Raudenbush & Bryk, 2002). For two-level models the ICC is calculated by:

$$ICC = \frac{\tau_{00}}{\sigma^2 + \tau_{00}},$$

where σ^2 is the level one variance and τ_{00} is the level two variance.

Applied researchers should have an estimate of the ICC for every nested level in their analyses because ignoring even relatively small amounts of interdependence (ICC values as small as .005) is known to inflate Type I error rate in single-level models (Mass & Hox, 2005; Raudenbush & Bryk, 2002). However, ICCs greater than .005 are commonly observed in hierarchical data structures. For example, the ICCs in a recent study of school climate on students' academic achievement range from .04 to .08 (Maxwell et al., 2017). ICCs as great as .522 have been observed in a study of players from track and field clubs (Swierzy et al., 2018). These examples of hierarchical data structures contained 2 levels of nesting. The formula for calculating the ICC is easily adapted for additional levels (more than two) by adding a term for the variance associated with each higher level to the denominator and inserting variance of the level of interest into the numerator and provides a point estimate of the proportion of variance in an outcome variable associated with a given level.

The Problem with Point Estimates

Point estimates, such as the ICC, have been criticized for being too dependent on the characteristics of a single, usually small, sample (Schmidt, 1996). This is especially problematic for multi-level modeling techniques that require relatively large sample sizes. As a result, the ICC may be significantly different from zero simply because of the large sample sizes commonly collected from hierarchical data structures. Furthermore, no informative guidelines exist for interpreting the magnitude of the ICC and the definition of a “meaningful” ICC depends on the

context of the research (Raykov et al., 2016). Some nationally recognized scientific organizations, such as the American Psychological Association, encourage researchers to report confidence intervals for each statistic because of these known problems with point estimates and null hypothesis significance tests in general.

Interval Estimation of the Proportion of Second-level Variance

Recent methods have been developed utilizing an interval estimation procedure to estimate the proportion of variance in an outcome variable attributed to higher levels of nesting that avoid the problem of relying on a null hypothesis significance test for the ICC. One such method is the CI-PSLV (Raykov et al., 2016). To determine the CI-PSLV, first an unconditional model is fit to the data to furnish an estimate of the standard error associated with the PSLV using the delta method (Raykov & Marcoulides, 2004). Second, the PSLV is calculated using the following formula:

$$PSLV = \frac{\tau_{\pi}}{\sigma^2 + \tau_{\pi} + \tau_{\beta}},$$

where σ^2 is the level one variance, τ_{π} is the level two variance, and τ_{β} is the level three variance. Finally, a 95% confidence interval is obtained based on the estimate of the PSLV and its standard error (Raykov et al., 2016). Please refer to Appendix A for an R-function that computes the endpoints of the CI-PSLV.

Rather than relying solely upon a point estimate, the CI-PSLV provides a range of plausible values for the PSLV in a population under consideration and is an informative supplement to the ICC (Raykov et al., 2016). The current study seeks to establish the accuracy of a 95% confidence interval of the PSLV in various sample sizes and under various degrees of second level dependence.

Method

The current study examines the accuracy of the CI-PSLV using simulated data. Data were generated in R 3.4.3 (R Core Team, 2017) using the following three-level model:

$$\text{Level 1: } Y_{ijk} = \beta_{0jk} + e_{ijk}$$

$$\text{Level 2: } \beta_{0jk} = \gamma_{00k} + r_{0jk}$$

$$\text{Level 3: } \gamma_{00k} = \delta_{000} + u_{00k}$$

$$\text{Composite: } Y_{ijk} = \delta_{000} + u_{00k} + r_{0jk} + e_{ijk} ,$$

where Y_{ijk} is the score of an individual i in second level j within the third level k on the dependent variable in the simulated data. δ_{000} is the fixed effect and the remaining terms represent random effects at levels three, two and one respectively in the composite model. The variance of Y_{ijk} is normally distributed.

Simulation Conditions

The following three conditions, as specified in Mass and Hox (2005), varied in the simulation: level two ICC (ICC = .1, .2 & .3), number of clusters in level two (L2NC = 30, 50 & 100), and number of clusters in level three (L3NC = 30, 50 & 100). The variance of Y_{ijk} was fixed to one in level one and .1 in level three across conditions, whereas the variance of Y_{ijk} at level two differed between conditions to vary the level two ICC. Sample size varied between conditions based on the total number of clusters in levels two (L2NC) and three (L3NC). Level one group size was fixed at 30 across conditions because thirty level-one units is a reasonable number to expect in educational settings (Mass & Hox, 2005). Level two ICC, L2NC and L3NC varied across 27 conditions (3 x 3 x 3).

Procedure

One thousand datasets were generated for each combination of conditions. An unconditional, three-level model was fit to each simulated dataset to demonstrate the accuracy of the CI-PSLV associated with a predicted outcome variable Y_{ijk} using *Mplus*. A model constraint was used to estimate the PSLV its standard error in all datasets. Please refer to Appendix B for *Mplus* source code for the estimation of PSLV and its standard error from the data generated with R. Estimates of the PSLV and their associated standard errors were saved for each dataset within each condition and were imported into R to calculate their 95% confidence intervals.

The accuracy of the CI-PSLV was assessed by comparing average standard errors of estimated PSLV, the proportion of CI-PSLV that include the actual ICC, and average width of CI-PSLV between simulated conditions. Accuracy of the CI-PSLV was assessed by the proportion of confidence intervals that included the population's level two ICC within a given simulation condition.

Results

Results from the simulation provide evidence in favor of the utility of the CI-PSLV under the conditions studied and are provided in Table 1. As can be seen from Table 1, the population level two ICC fell within the CI-PSLV in about 95% of repeated samples. Standard errors associated with the estimate of PSLV and the average width of confidence intervals tended to decrease as sample size (L2NC & L3NC) increased and were slightly larger in conditions with larger ICC conditions. The CI-PSLV seemed to capture the actual second level ICC and none of the confidence intervals included zero. Lower standard errors of the estimate of PSLV were associated with a greater number of level three clusters relative to level two clusters in conditions with equal, overall sample size (L3NC=100 & L2NC = 50 compared to L3NC=50 & L2NC=100;

L3NC=100 & L2NC = 30 compared to L3NC=30 & L2NC=100; and L3NC=50 & L2NC = 30 compared to L3NC=30 & L2NC=100).

Discussion

The current study provides evidence of the applied utility of the CI-PSLV. The CI-PSLV appears to capture the actual second level ICC as long as sample size requirements for multi-level modeling are met to begin with as described in Mass and Hox (2005). None of the confidence intervals in the current study included zero. Future work needs to be conducted to determine the usefulness of the CI-PSLV for detecting a second level ICC close to zero (smaller than .1 which is the lower bound of the current study). Similarly, research needs to be conducted to address the robustness of the CI-PSLV to violations of the assumption that outcome measures (Y_{ijk}) are normally distributed.

Multi-level modeling is appropriate in a variety of fields given the inherent, hierarchical nature of data they utilize. Similarly, the application of an interval estimate of the PSLV is appropriate anytime three-level data structures are encountered, regardless of the specific research area. The results of the current study provide evidence of the accuracy of the CI-PSLV, which avoids many of the known problems associated with relying solely on point estimates for null hypothesis significance tests of the traditional ICC. This does not suggest the CI-PSLV should replace the traditional ICC, but rather, it should be reported as another piece of evidence in conjunction with the ICC (Raykov et al., 2016). Once its efficiency is established, the CI-PSLV procedure will provide a range of plausible estimates for the amount of interdependency of scores due to hierarchical data structures and that should be reported in addition to the traditional ICC.

References

- Maas, C. J. M., & Hox, J. J. (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica*, 58, 127-137.
- Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, 1, 86-92.
- Maxwell, S., Reynolds, K. J., Lee, E., Subasic, E., & Bromhead, D. (2017). The impact of school climate and school identification on academic achievement: Multilevel Modeling with student and teacher data. *Frontiers in Psychology*, 8, 1-21.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models. Application and data analysis methods* (2nd ed). Thousand Oaks, CA: Sage.
- Raykov, T. (2010). Proportion of third-level variation in multi-level studies: A note on an interval estimation procedure. *British Journal of Mathematical and Statistical Psychology*, 63, 417-426.
- Raykov, T., & Marcoulides, G. A. (2004). Using the delta method for approximate interval estimation of parameter functions in SEM. *Structural Equation Modeling*, 11, 621-637.
- Raykov, T., Patelis, T., Marcoulides, G. A., & Lee, C. (2016). Examining intermediate omitted levels in hierarchical designs via latent variable modeling. *Structural Equation Modeling*, 23, 111-115.
- R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>

- Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods, 1*, 115-129.
- Swierzy, P., Wicker, P., & Breuer, C. (2018). Usefulness of multilevel modeling in sport management research: The case of voluntary roles in nonprofit sports clubs. *Measurement in Physical Education and Exercise Science, 22*, 1-12.

Table 1.

Average Standard Errors of the PSLV Estimate, Number of Intervals Capturing ICC & CI
Width Across Simulation Conditions for Data Generated Using R

		<u>Level 2 ICC</u>								
Number of Clusters		.3			.2			.1		
L3NC	L2NC	S.E.	#CI	Widt	S.E.	#CI	Widt	S.E.	#CI	Widt
				h			h			h
100	100	.0043	954	.0169	.0033	957	.0131	.0020	942	.0080
	50	.0055	957	.0216	.0043	952	.0168	.0026	954	.0104
	30	.0068	950	.0268	.0053	954	.0209	.0034	952	.0129
50	100	.0061	951	.0239	.0047	951	.0186	.0029	956	.0113
	50	.0078	951	.0305	.0061	951	.0238	.0037	949	.0146
	30	.0097	957	.0379	.0075	954	.0295	.0047	951	.0183
30	100	.0079	956	.0309	.0061	960	.0240	.0037	962	.0146
	50	.0101	959	.0394	.0078	955	.0307	.0048	956	.0189
	30	.0125	940	.0489	.0097	940	.0381	.0060	940	.0237

Appendix A

R-Function for Interval Estimation of Proportion of Second-Level Variance from Raykov et al.
(2016)

```
ci.pslv = function(pslv, se){  
  l = log(pslv/(1 - pslv))  
  sel = se/(pslv*(1 - pslv))  
  ci_l_lo = l-1.96*sel  
  ci_l_up = l+1.96*sel  
  ci_pslv_lo = 1/(1+exp(-ci_l_lo))  
  ci_pslv_up = 1/(1+exp(-ci_l_up))  
  ci = c(ci_pslv_lo, ci_pslv_up)  
  ci  
}
```

Appendix B

Mplus Source Code for Estimation of the Proportion of Second-Level Variance adapted from
Raykov et al. (2016)

TITLE: Interval Estimation of PSLV from Simulated Data

DATA:

FILE IS Mpluslist.txt;

TYPE=MONTECARLO;

VARIABLE:

NAMES ARE L1ID L2ID L3ID repID u00k r0jk eijk y;

USEVARIABLE ARE y;

CLUSTER ARE L3ID L2ID;

ANALYSIS:

TYPE = THREELEVEL;

ESTIMATOR IS ML;

MODEL:

% WITHIN%

y* (P1);

%BETWEEN L2ID%

y* (P2);

%BETWEEN L3ID%

y* (P3);

MODEL CONSTRAINT:

NEW(PSLV);

PSLV=P2/(P1+P2+P3);

OUTPUT:

STDYX;

SAVEDATA: RESULTS = results.dat;

The History, Evolution, and Trends of Academic Dishonesty: A Literature Review

Amy Zachek

Department of Educational Administration

University of Nebraska-Lincoln

Abstract

Academic dishonesty is a murky problem without a commonly agreed upon solution in American higher education. It has a long-standing history in higher education but a short history in academic literature, it has evolved rapidly and longitudinally (McCabe & Trevino, 1996), and it has several easily apparent trends and others that the majority of researchers are in disagreement about. While traversing this perilous landscape of dichotomies, this paper will examine connections and gaps in the literature, make suggestions and recommendations for future study based off of these results, and examine the implications that these recommendations could have on higher education policy. Results demonstrate that a learning-based, non-penal approach to academic dishonesty may be the most helpful stance that institutions can take for their students. Key areas of interest to the author include the intersections of technology, generational change, and self-authorship, and these areas will be focused on in detail throughout the literature review.

Keywords: college students, higher education, cheating, academic dishonesty, history, evolution, trends

doi: 10.32873/unl.dc.ne006

In the realm of higher education, there is one cardinal sin above all others: academic dishonesty. After all, the theft of thought is antithetical to the entire institutional purpose of education. In reviewing the literature surrounding academic dishonesty, I will explore its historical context and the history of its presence in the literature, trace its evolution, and analyze contemporary trends. In the course of this chronological review, I will explore several pieces that are important to a richer understanding of this topic, including demographics, generational change, and technology. In this way, I will synthesize the history of academic dishonesty, how it has evolved, and its contemporary status in American higher education, and use that analysis to discover gaps, challenges, and implications for the future.

Throughout the course of this literature review, I will utilize the terms academic dishonesty and cheating. I define academic dishonesty as any act of cheating in isolation, collusion with other students, or plagiarism. Allemand (2012) broadly defines academic dishonesty as any sort of unfair academic advantage, and I agree with the spirit of this broad definition. For the purposes of this literature review, I consider academic dishonesty and cheating to be analogous terms and I will refer to the two interchangeably.

History of Academic Dishonesty

An immediate assumption one might make is that academic dishonesty is a recent phenomenon; an epidemic of modern society that did not affect our more morally austere predecessors. Chace (2012) illustrates that this is far from the case. He writes, “Indeed, every study over the decades has concluded that cheating at American colleges and universities is rampant” (Chace, 2012, p. 23). Ergo, cheating is not a problem that has cropped up in recent years; studies simply reaffirm time and time again that it exists. Echoing this sentiment, Arnold et al. (2007) write “Since the beginning of formalized education, student cheating has been a

challenge for educators” (p. 2). In Blum’s (2009) book “My Word! Plagiarism and College Culture,” she delineates the colorful and varied history of cheating around the world. One of the oldest examples comes from China, where there was a 1400 year history of cheating for the keju exam for civil service (Blum, 2009). A startling example from the history of American higher education comes from Yale, when “cheating scandals involved half the student body at Yale in the 1860s” (Blum, 2009, p. 25). From these examples, it’s easy to see that academic dishonesty is not a new problem.

Literature Chronology

Even with the lengthy history of academic dishonesty in education, the articles that I found for this review seem to suggest that literature in this area first originated in the twentieth century. Several authors consider Bowers’ (1964) work to be the seminal, landmark academic study of academic dishonesty (Chace, 2012; McCabe & Trevino, 1996). Chace (2012) writes that Bowers’ work was the “first comprehensive study of cheating at colleges and universities” (p. 23). However, though Bowers’ work may have been the first influential work in the field, it was by no means the first study of academic dishonesty. Bowers’ (1964) study identifies several precursors to his research in the field, including a 1920s study about the causes of dishonesty, a 1935 study about honor systems, and the 1960 Cornell Value Study. The impressive thing about Bowers’ (1964) study is that its scale was unprecedented in the field and that its findings are important enough to still be relevant to the field over 50 years after it was published.

One of Bowers’ (1964) primary conclusions is that, at least according to one set of questions regarding academically dishonest behaviors, “fully three-quarters of the students have engaged in at least one act of academic dishonesty; half have committed two or more” (p. 48). Modern academic dishonesty does not look much different, and at least two studies since 1964

have illustrated that cheating has only trended upward since the Bowers study (McCabe et al., 1996; McCabe & Trevino, 2001). McCabe et al. (2001) state that “Cheating is prevalent and some forms of cheating have increased dramatically in the last 30 years” (p. 219). McCabe and Trevino (1996) don’t necessarily claim a sharp uptick in cheating, but they describe a small and consistent increase since Bowers’ cheating study (McCabe & Trevino, 1996; Bowers, 1964). Recent works such as Tavis and Aronson’s (2020) piece have gone beyond a demographic study of academic dishonesty and have attempted to explain why it may be easier for some students to participate in cheating behaviors than others. Tavis & Aronson (2020) posit that students who have made the decision to cheat (or not cheat) are more resolute in their decision and belief system after they have made that choice, and that each future decision makes them feel even more resolute in their personal decisions. More future studies that examine the reasons why students cheat will be necessary to help provide insight on this topic.

Evolution of Academic Dishonesty

Over the last half century, one of the key ways that academic dishonesty has evolved is by way of changed generational attitudes. Blum (2009) does an excellent job describing some of the generational shifts which have created an ideal cultural temperature for academic dishonesty in academia today. She writes about the social shift from valuing an authentic self to valuing a performance self, and how that has had a critical impact on youth culture (Blum, 2009). Blum (2009) states, “If in the past a stable and integrated identity was regarded as the basis of mental health, in recent years a fluid, flexible self is seen as more responsive and more socially desirable” (p. 79). She then relates this idea to the traditional academic concepts of authorship and intellectual property, writing, “This self, made out of various readily available components, is as unconcerned about strict attribution of authorship as it is about strictly tracing an

unchanging identity,” (Blum, 2009, p. 79). In this way, this flexible self may have a greater impact on students’ propensity to cheat than higher education administrators may realize (Blum, 2009)

Technology

This fluid identity may have partially come about due to modern technology, and the common concept of “identity play” in online gaming (Blum, 2009, p. 78). So, while online gaming doesn’t directly interact with academic dishonesty, the effect that it may be having on student attitudes *does*. Hamlen’s (2012) piece about video gameplay also discusses these changes in attitudes. Hamlen (2012) writes, “This study was based on the theory that current media-immersed children may think and learn differently than those in the past,” (p. 1146). Though her results do not find strong evidence for her theory about video games affecting academic dishonesty, she does claim that “new technologies are enabling new ways of learning and thinking” (Hamlen, 2012, p. 1146).

When it comes to technology and cheating, the literature mainly seems to focus on the technology that assists cheating detection and not how students use technology to cheat. For example, Sledge and Pringle (2010) mention a few websites that help instructors fight cheating with technology, stating “In college, technology has allowed cheaters to be quickly identified through websites such as turnitin.com and essayrater.com” (p. 4-5). Blum (2009) goes into more detail about how students use technology to cheat. She writes about Wikipedia, but chooses to focus on how this is another platform which is shaping generational attitudes about cheating instead of the technological side of how students use it to cheat (Blum, 2009). She writes, “Wikipedia is one technological innovation that takes to its logical conclusion the general unconcern among the young about tracing individual contributions to a written product. This is

regarded with great consternation by educational authorities, guided by their model of the authentic self” (Blum, 2009, p. 70). In this way, we can see that the intersection of ever-present technology and generational attitudes has had a powerful impact on the evolution of academic dishonesty.

Trends in Academic Dishonesty

Us Vs. Them

A key trend in the literature is the concept of the students vs. teachers or us vs. them (Helgeson, 2002; Wideman, 2011). One student in Helgeson’s (2002) study incorporates this concept into their response about how faculty handle cheating, stating, “It’s kind of like ‘students vs. teachers’ and we help each other out” (p. 124). With this divide between students and teachers, it makes sense that there are also correspondingly differing opinions about academic integrity policies from both students and faculty (Helgeson, 2002; Allemand, 2012). It’s clear that finding a universal solution that all stakeholders agree upon is one challenge. Hamlen (2012) even goes so far as to suggest that the “disparity between student and teacher beliefs about what constitutes cheating” (p. 1146) is one of the main reasons for high rates of cheating. She also writes about nature versus nurture as initial factors that lead to cheating behaviors. Hamlen (2012) states that there are differing opinions in the field about whether environmental factors or the personality of the student is the cause of cheating.

Honor Codes

Arnold et al.’s (2007) piece about honor codes echoed Bowers’ (1964) sentiments about climate by stating that “a culture of academic integrity may be the single most determining factor regarding student cheating on a college campus” (p. 3). Arnold et al. (2007) differ with Bowers (1964) in finding that there are not significantly lower amounts of cheating at schools with honor

codes. Bowers (1964), on the other hand, writes that “Schools with honor systems have by far the lowest rate of academic dishonesty” (p. 184). It’s important to note that Bowers’ (1964) data and opinion on this subject is over fifty years old and could be outdated due to the different historical contexts in the period. For example, Bowers (1964) writes that honor systems were especially prevalent at single-sex schools, and it’s possible that many of those institutions might have closed or substantially changed in the second half of the 20th century due to the Civil Rights Act (1964). Regardless of the efficacy of honor codes, Arnold et al. (2007) show that these systems have returned to higher education in force in a (well-intentioned but potentially misguided) attempt to fight the prevalence of academic dishonesty.

Social Changes

Next, Blum’s (2009) piece provides an anthropological look at student life and youth culture, and how that affects their outlook on academic integrity. Taking the stance of an ethnographer, Blum (2009) points out a number of social trends that she has observed. These observations include that the Generation Y and Z enjoy informal intertextuality, that students have been raised to be sociable, and that society has transitioned from valuing the authentic self to valuing the performance self (Blum, 2009). She describes that these attitudes of the younger generations are what have led to a “mismatch between the academy’s and student’s expectations” (Blum, 2009, p. 179). In this way, she shows that the new performance self directly contradicts age old ideas of authorship and intellectual property (Blum, 2009). Through these points, Blum (2009) illustrates social change as a trend in the recent years of academic dishonesty. She would likely suggest that academe is increasingly out of touch with a youth culture that values intertextuality, sharing, and communal authorship.

An example of this cultural discrepancy can be found in Chace's (2012) essay. He states that "the essay's fingerprint" (Chace, 2012, p. 23) or its originality is the goal of academic writing. This thought process is in line with the old guard of the academy when one contrasts it against Blum's (2009) idea of the authentic self and authorship as irrelevant to the current generation. Interestingly, he is aware of the double-standards that today's youth experience with intertextuality and academic citing standards. Chace (2012) writes, "a wrong in one place is highly prized in another" (p. 26), a point which connects back with Blum's (2009) suggestion about everything relating to cultural relativism and shifted generational attitudes. Beasley (2016) also echoes these ideas in stating that in youth culture "sharing and quoting without attribution are established norms of conduct" (p. 56). He takes this a step further and writes that, "In addition to this hypersharing reality that many of today's college students live in, the students are also embedded in an overarching American culture where honesty does not seem very valued, especially when being dishonest can get one "ahead,"" (Beasley, 2016, p. 56). After all, if those are the norms that we are drilling into our youth, naturally they are going to demonstrate that societal training in their writing and work.

External and Internal Values

Another significant trend is the importance of the external value system on campus and students' internal values. Bowers (1964) writes "The combination of weak disapproval of cheating and poor grades results in more than four out of five students' being cheaters" (p. 78). He also uncovers the importance of internal values on cheating. If parents put importance on good grades, students are more likely to cheat; if students put more importance on good grades, they are less likely to cheat (Bowers, 1964). In other words, Bowers (1964) is suggesting that

the desire to have good grades and to succeed must come from within; if that pressure comes from the outside, students may strive to achieve the grade at any cost to receive external praise.

Demographics

Beasley's (2016) article takes a look at the demographics of students who are reported for academic dishonesty. In line with the other research (Allemand, 2012), he writes that, "few official sanctions levied upon students for academic dishonesty" (Beasley, 2016, p. 45). In looking at the demographics of who cheats, one of Beasley's (2016) primary concerns is if students of minoritized racial backgrounds are more likely to be reported due to institutionalized racism. To this end, Beasley (2009) reports that, "International undergraduates were much more likely to get reported for academic dishonesty than were domestic students" (p. 55). At this time, there does not seem to be a clear answer for this difficult question, but this is an important area for future critical research.

When it comes to the demographics of gender, two studies are in agreeance that women students cheat less than men students (Bowers, 1964; Hamlen, 2012; Helgeson, 2002). However, these views are widely contested in other studies. Arnold et al. (2007) and Beasley (2016) find no gender differences when it comes to cheating, the latter also echoes the inconsistencies of findings and beliefs regarding gender and cheating in the literature. Beasley (2016) does bring up an interesting idea that originates from McCabe, which is that "the previously found gender differences in cheating behavior have diminished to being virtually nonexistent in the second decade of the 21st century" (p. 49) and that that accounts for one explanation of the inconsistency with previous research findings.

Many other demographics are at play besides gender. Once again, most of the literature seems to proffer its own views on the demographic characteristics of a student who cheats.

Beasley (2009) lists the following as characteristics of students who are often caught cheating:

The lower the GPA, the more likely it was that the student would be found guilty. In addition, non-Whites, students in fraternities or sororities, transfer students, and students who were not varsity athletes were more likely than their contemporaries to be found guilty. (p. 46)

Looking at the demographic of major, Sledge and Pringle (2010) find that there is “disproportional cheating by business students” (p. 5). More than anything, these disparate views show that students of any intersection of identities can commit an act of academic dishonesty.

Gaps in the Literature

Now that we have explored the history, evolution, and trends of academic dishonesty, let’s look into the gaps in the literature. One self-admitted gap of Bowers’ (1964) study is that it would be better to compare intellect ability to cheating, not just grades (GPA). He writes, “Instead, we must make do with the student’s grades in college, which reflect not only his ability but also his motivation” (Bowers, 1964, p. 73). Another gap is the issue of self-reporting, which is a perennial issue for most all of the studies included (Blum, 2009; Sledge & Pringle, 2010). To this end, Blum (2009) writes that “we can understand the students to be not exactly “telling the truth” but rather performing a virtuous – or daring – self in front of a peer interviewer” (p. 8).

Like Arnold et al. (2007), Sledge and Pringle (2010) also look at honor codes in their piece. Rather than just assessing the landscape of academic dishonesty, Sledge and Pringle (2010) take a slightly different approach by applying a set of interventions at their institution,

studying their effect, and then applying a second set of interventions to try to create broad cultural change. In the end, they conclude that “students may be the best resource to teach their peers about the honor code” (Sledge & Pringle, 2010, p. 9). Another gap is that their study was only held at their campus’ business college. A broader approach would be a better application of the efficacy of honor code interventions, particularly in combination with some of their recommendations of connecting with high school students and seeking college student buy-in up front (Sledge & Pringle, 2010).

One gap in Chace’s (2012) piece is that he does not seem to be familiar with the literature that shows that cheating happens in nearly equivalent amounts at colleges with and without honor codes (Arnold et al., 2007). Instead, he enshrines schools with honor codes as saviors of academic integrity, writing, “They see the dangers of cheating for what they are: practices in which many students can be hurt by the dishonesty of a few” (Chace, 2012, p. 31). Higher education administrations should take care to beware of wearing rose-colored glasses and to avoid an over-romanization of honor codes as an academic panacea.

Another gap, as pointed out by Beasley (2009), is the potential effect of racism on the reporting of students for academic dishonesty. A qualitative study with a critical theory approach on the effect of stereotypes and racist beliefs would be prudent. A parallel gap to this is research that explores if environmental factors or the personality of the student who cheats is a bigger cause of cheating behaviors. As explored earlier, Hamlen (2012) illustrates that there are differing opinions in the field on the subject, so this is definitely an area where additional research would help shed light on the subject. More research about the motivations for academic dishonesty in general would help inform academic dishonesty policy and ensure that institutional interventions are getting at the root cause of cheating on campus.

Some consider differences in cheating in relation to gender to be outdated in our society where gender norms do not matter as much as they once did (Beasley, 2016). While it is true that gender norms have changed in American society, the literature I reviewed made no mention of cheating in relation to students who don't identify with the normative gender binary. Non-binary, transgender, and other students were not even categorically mentioned in the literature I reviewed. This is certainly another gap and could also be a potential area for future inquiry, since I did not find literature to exist yet about academic dishonesty for this specific student population.

A final area for potential future study comes from a postulation in Bowers' (1964) study. He suggests that one possible ill-effect of academic dishonesty is that it may lead to professionals in the workplace who disregard ethics (Bowers, 1964). One potential recommendation for future research could be a study of ethics in the workplace for college cheaters vs. non-cheaters. Bowers (1964) already takes steps to analyze high school cheaters and how that affects cheating in college, and discovers that, "those most prone to cheat [in college] are those who cheated in high school and have been consistently socially oriented" (p. 137). It would be interesting to take this idea a step further and look at the correlation between college cheating and the workplace. Similarly to high school cheaters being more likely to cheat in college, are college cheaters more likely to continue their behaviors and to "cheat" (e.g. disregard ethics and the rules, do things out of personal gain, etc.) in the workplace? This current gap in the researcher would show the real-world importance of effectively intervening with academic dishonesty before students enter the workplace.

Recommendations

It's clear from the literature that there are many divisive opinions about academic dishonesty, and that it's not a subject with clean-cut answers. Therefore, my two key recommendations are that more research is needed about student motivations and reasons for academic dishonesty and that an educational approach to responding to academic dishonesty is optimal. For the first recommendation, colleges and universities will never be able to appropriately respond to and stop academic dishonesty until they understand where their students are coming from (Kibler & Kibler, 1993). In this way, I consider the heart of the academic dishonesty problem to be an issue of understanding which requires more research, educational initiatives, and dialogues between students and administrators.

Blum (2009) writes that academic integrity should not be viewed as an issue of morality or criminality. Instead, she proposes that "a third approach treats academic integrity, especially the mandate to cite sources, as a set of academic skills to be learned" (Blum, 2009, p. 165). In other words, frame this as a learning opportunity. Treat academic integrity like any other academic subject and teach students *why* it matters. This is also where researching and understanding more about student experiences, perceptions, and motivations of academic dishonesty will show results. Administrators will start to understand why cheating is happening in the first place and to create new programs to target these areas of student development. Students will understand why academic integrity is important by receiving formalized instruction about it, and administrators will be able understand the foundational reasons why students cheat and to provide the support that students need to eliminate cheating as a coping strategy. I postulate that this two-fold approach could be one way to approach solving the academic dishonesty problem on campus.

Conclusion

In this literature review, I have synthesized the history and contemporary status of academic dishonesty in American higher education. I hope to have revealed patterns, gaps, and potential recommendations for higher education administrators and future study. The lenses of demographics, generational change, and technology provided helpful insights along the way, and I think that the intersection of each of these areas with academic dishonesty are their own rich areas of study. This field is particularly ripe for study since it is still a murky subject for colleges and universities without universally agreed upon solutions. One potential way to get to that point is to study the problem in greater detail, to create educational opportunities for students to learn more about how to create academic work with integrity, and to create opportunities for administrators to learn more about why students cheat. After all, in describing one example of a multiple offense cheater, Blum (2009) writes that the student “really wanted her to *learn how* to do this work right” (p. 175). And isn’t helping students get to that point the mission of education in the first place?

References

- Allemand, K. (2012). Student and Faculty Perceptions of Academic Dishonesty: A Qualitative Single-Case Study. Northcentral University.
- Arnold, R., Martin, B. N., & Bigby, L. (2007). Is there a relationship between honor codes and academic dishonesty? *Journal of College and Character*, 8(2), 1-20.
- Beasley, E.M. (2016) Comparing the demographics of students reported for academic dishonesty to those of the overall student population. *Ethics & Behavior* 26(1), 45-62.
- Blum, S.D. (2009) *My Word! Plagiarism and College Culture*. Cornell University Press.
- Bowers, W. J. (1964). *Student dishonesty and its control in college*. Bureau of Applied Social Research, Columbia University.
- Chace, W. M. (2012). A question of honor. *American Scholar*, 81(2), 20–32.
- Hamlen, K. R. (2012). Academic dishonesty and video game play: Is new media use changing conceptions of cheating? *Computers & Education*, 59(4), 1145–1152.
- Helgeson, K. M. (2002). *Perceptions of academic dishonesty among students and faculty at the University of Nebraska-Lincoln*. University of Nebraska--Lincoln.
- Kibler, W. L., & Kibler, P. V. (1993). When Students Resort to Cheating. *The Chronicle of Higher Education*. <https://www.chronicle.com/article/When-Students-Resort-to/70647>
- McCabe, D. L., Trevino, L. K., & Butterfield, K. D. (2001). Cheating in Academic Institutions: A Decade of Research. *Ethics & Behavior*, 11(3), 219–232.
- McCabe, D. L., & Trevino, L. K. (1996). What We Know About Cheating In College *Longitudinal Trends and Recent Developments*. *Change: The Magazine of Higher Learning*, 28(1), 28–33. <https://doi.org/10.1080/00091383.1996.10544253>
- Sledge, S., & Pringle, P. (2010). Assessing honor code effectiveness: Results of a multipronged approach from a five year study. *Research & Practice in Assessment*, 5.
- Tavris, C., & Aronson, E. (2020). *Mistakes were made (but not by me) third edition: Why we justify foolish beliefs, bad decisions, and hurtful acts*. Houghton Mifflin Harcourt.

Wideman, M. (2011). Caring or collusion? Academic dishonesty in a school of nursing.
Canadian Journal of Higher Education, 41(2).

Bilingual Acculturation Assessment: An Overview of Current Developments

Donna Chen

Department of Educational Psychology
University of Nebraska-Lincoln

Abstract

Assessments are often used for decision-making in education, mental health practices, and industry. Consequently, decisions based on these assessments affect multiple aspects of a person's life. Given the increase of ethnic minorities in the U.S., factors concerning the appropriateness and interpretation of tests based on norms must be reconsidered. The multifaceted effects of culture are just one factor to consider so as to not overlook important cultural components that may negatively impact the decision-making process. Additionally, language, with close ties to culture, must also be considered. Thus, the complexity of culture and language in tandem to assessment-based decision-making necessitates fundamental understanding of the effects of bilingual acculturation assessment development. A brief history of bilingual acculturation is considered. Careful consideration is given to bilingual acculturation assessment within the context of measurement development, empirical application, and validity issues. Implications and limitations of the current studies are discussed.

Keywords: bilingual acculturation assessment, measurement development, ethnic minorities, theoretical/conceptual

doi: 10.32873/unl.dc.ne007

Bilingual acculturation assessment: An overview of current developments

Culture is both a messy and a beautiful phenomenon; messy in the sense that it is not easily defined, yet simultaneously beautiful in that it creates a tapestry of rich diversity, woven in to the human race. Culture touches all facets of life. It governs how we speak, the ways we behave, our small mannerisms, to the way we interact with each other. Culture is not a stagnant entity, rather, it shifts dynamically, often more apparent when people of different backgrounds consistently interact with one another. As the ethnic minority populations in the United States continue to increase, currently forming about 23.5% of the population (U.S. Census Bureau, 2018), people of different cultures come into continuous contact with one another, thereby changing the other's culture through a process known as acculturation—the extent to which a person associates with a given ethnic minority or majority culture (Moyerman & Forman, 1992).

Given the increase of ethnic minorities in the U.S., factors concerning the appropriateness and interpretation of tests based on norms must be reconsidered, especially when these tests are involved in decision-making (Sandoval, 1998). Norms can be conceptualized as an aggregated score distribution derived and standardized from a large representative sample against which individual performance can be compared. The norms for these tests, often measured and used as a reflection of the *average* test taker in the population (Sandoval, 1998), may often be based on the majority-group. The implication for drawing from the majority-group is that test scores may not be reflective of individuals outside the majority group. These tests are often used for decision-making in education, mental health practices, and industry, affecting multiple facets of a person's life. Thus, cultural norms must be taken in to account so as to not overlook important cultural components that may negatively impact the decision-making process (Sandoval, 1998).

An important factor closely tied to cultural norms is acculturation. It is important to consider how a person's acculturation status may affect their performance on tests.

Language is closely tied with culture and must be considered when measuring acculturation (Luna et al., 2008; Marín & Gamba, 1996; Schwartz et al., 2014). Bicultural bilinguals, individuals who associate themselves with two cultures and speak both languages, have been found to have “distinct cognitive frameworks” when speaking different languages (Luna et al., 2008, p. 279). This phenomenon of switching between frameworks depending on which language is spoken is often referred to as cultural frame-switching (Benet-Martínez et al., 2002; Luna et al., 2008; Schwartz et al., 2014). Thus, the language of a test or assessment may activate certain characteristics, thereby differentiating a bilingual person's results had they taken the test in the other language. It is important to consider how bilingual acculturation assessment is conducted, as it has implications for decision-making when using normed tests.

A Brief History of Bilingual and Bicultural Assessment

In considering why acculturation measures are so important, two cases are presented that served as critical points in drawing public attention towards consideration of equitable education for culturally, ethnically, and linguistically diverse students. The first case was the *Diana v. California State Board of Education* (1970) court case. Nine Mexican-American children whose primary language was Spanish were given IQ tests in English and subsequently placed in Special Education classes due to their low scores. The court ruled that California school districts were to test children in their primary language as well as with nonverbal tests (McLean, 1995). Second, the *Lau v. Nichols* (1974) U.S. Supreme Court case concerned about 1,800 Chinese students in California who were placed into separate “Oriental” English only schools, subsequently to be

ignored by teachers and school administrators (McLean, 1995). The court ruled that schools must provide equal education to all students, regardless of English language ability.

Although the cases put into ruling what schools should do, positive results adhering to the *intentions* of the ruling were not immediate. Cummins (1982) reported the continuous disproportionate number of bilingual students who are labeled “low IQ” and placed in special education classes. He notes the bias of applying mental tests that fail to account for cultural and linguistic differences. Additionally, he illustrates the need for school programs to be culturally and linguistically oriented, to include context-embedded approaches towards achieving language proficiency. With a more specific focus on school psychologists’ use of testing, Figueroa (1989) brings forth the issue of anomalies in test scores among bilingual students. Again, there is evidence to suggest that, since professionally developed tests are often normed based on norm groups consistent with the United States census, consequences of using such tests on non-norm populations may lead to errors and misdiagnoses. Additionally, norming based on census data does not provide insight into cultural differences within groups. Figueroa (1989) makes the case for the development of appropriate linguistically and culturally-relevant tests for use in schools.

Framework

To begin, the framework for this paper is built by discussing pertinent key concepts along with their associated issues.

Culture

First, as discussed earlier, culture is not easily defined. Many social and behavioral scientists have attempted to define culture, each covering slightly different aspects. As Frisby (1998) summarizes, culture has been defined to include personality (Kroeber & Kluckhohn, 1952); shared learning across generations (Rohner, 1984); “the totality of ideals, beliefs, skills,

tools, customs, and institutions into which each member of a society is born” (Sue & Sue, 1990, p. 35); as well as “some combination of differences in skin color, country of origin, language, customs, socialization priorities, and sometimes socioeconomic class” (Triandis & Brislin, 1984, p. 1007). Indeed, the definition of culture for this paper is based on an amalgamation of these definitions, which becomes apparent throughout the paper, as it was found that different studies focus on different aspects of culture as a measure of acculturation.

Acculturation

The most often cited framework for understanding acculturation is John Berry’s (1980) four acculturation profiles: integration, assimilation, separation, and marginalization.

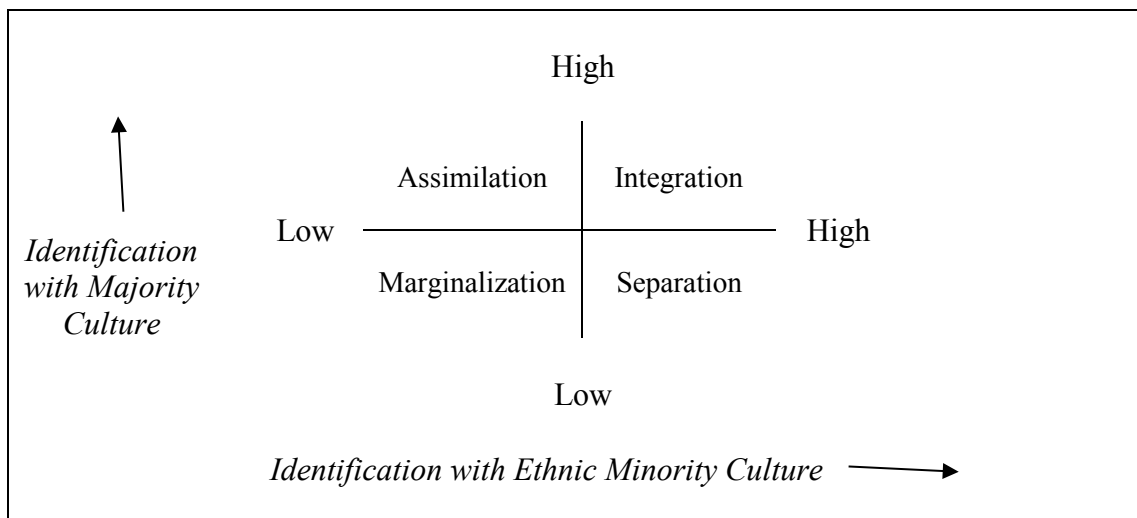


Figure 1. Visual display of Berry’s (1980) four acculturation profiles.

Each of these are considered from the basis of how much a person values and identifies with the ethnic group versus the dominant (or majority) group. Figure 1 is a visual display of how these four acculturation profiles relate to each other. Integration occurs when a person values both their ethnic and majority culture, finding a complementary view for having both, often referred

to as biculturalism. Assimilation is when the person values only the majority culture, having either a low or no value or identification with their ethnic culture. Separation, on the other hand, is the manifestation of only valuing the ethnic culture while disassociating with the majority culture. Marginalization is when the individual rejects the values of both the ethnic and majority cultures.

Follow-up studies with different ethnic minority groups have demonstrated different, or separate, acculturation groups from those originally proposed by Berry (1980). For example, in a study with Chinese Canadian university students, Chia and Costigan (2006) identified five acculturation groups instead of four. Schwartz and Zamboanga's (2008) study with Hispanic college students found six latent classes emerged from the analyses. This discrepancy in findings suggests that the true nature of acculturation cannot be easily pinned down and may even differ by ethnicity groups. Since acculturation measures are used for a variety of purposes (e.g., determining a student's educational placement or allotment of additional educational resources or services; determining mental health status; determining language of assessment, etc.), often having a major impact on people's lives, there is need to more carefully examine how acculturation is being measured.

Biculturalism

Biculturalism, or Berry's integration profile, is the endorsement of both the culture of origin as well as the receiving culture, often manifested as identification with both the ethnic minority and majority (Chen et al., 2008; Schwartz & Unger, 2010; Van de Vijver & Phalet, 2004). Initial attempts at measuring biculturalism included considering proficiency and comfort in both languages, choice in friends, media preferences, etc. (Schwartz & Unger, 2010). Progress in the area then led to a broader consideration of biculturalism to include the cultural practices

that an individual may partake in, the values from each culture that the individual adheres to, as well as which culture, or both, the individual identifies with (Schwartz & Unger, 2010). With different perspectives for this subject, Basilio et al. (2014) continued to extend biculturalism towards encompassing more psychological constructs such as identity, beliefs, attitudes, and response towards cultural cues.

Even within biculturalism, there are differences in how individuals perceive the compatibility and complementarity of their cultures—known as bicultural identity integration (BII; Benet-Martínez et al., 2002). While some bicultural individuals see their two cultures as being compatible and complementary, others may perceive their two cultures to contradict in their meaning systems. Benet-Martínez et al. (2002) conducted a study that revealed BII to be a moderator for the effects of cultural frame shifting—unconscious behavioral switches, or frame of thinking, primed by cultural cues—on measures of acculturation. In instances where the two cultures are perceived to be incongruent (e.g., individualistic versus collectivistic cultural values), the focus is on how bicultural individuals negotiate those oppositional values. When given certain cultural primes or cues, an individual is more or less likely to respond in a way that is congruent to the culture of the primes. Thus, assessments that ignore levels of acculturation face issues of validity.

Comas-Díaz and Grenier (2002) briefly outline some of the existing measures of acculturation, many of which are culture specific. Some examples include the Acculturation Rating Scale for Mexican Americans I (Cuéllar et al., 1980) and II (Cuéllar et al., 1995), the Behavioral acculturation Scale for Cubans (Szapocznik et al., 1978), the Suinn-Lew Asian Self-Identity Acculturation Scale (SL-ASIA; Suinn et al., 1987), and the Acculturation Scale for American Indian Adolescents (King & Keane, 1992). These examples illustrate that within some

of the largest ethnic minority groups (i.e., Hispanic/Latino, Asian), there are many sub-cultures that vary vastly. Each of these culture-specific scales make it difficult for researchers to generalize the results to other ethnic groups. There remains, however, the reality that creating an acculturation scale to fit all existing cultures is extremely difficult. Due to the dissonance between studies in minority ethnic groups, failure to account for acculturation status results in violations of reliability and validity (Padilla, 2001). Therefore, the question becomes, what work has been done to ameliorate these discrepancies? This paper will review developments in bilingual acculturation assessments, validity and reliability issues, as well implications for these considerations.

Methods

The search phrases bilingual assessment, acculturation assessment, and bilingual acculturation assessment were entered into various search engines and databases, including Google Scholar, Academic Search Premier, and PsycInfo (EBSCOhost). Articles were selected based on their relevance to the topic of this paper.

Results

The articles that were found can be categorized to two sections: measure development and empirical work.

Measure Development

Hispanic Ethnic Groups

The Acculturation Rating Scale for Mexican Americans (ARSMA; Cuéllar et al., 1980) was developed by giving a sample of Mexican Americans a measure in English and Spanish, scored on a 5-point Likert scale: Very Mexican, Mexican-Oriented Bicultural, True Bicultural, Anglo-Oriented Bicultural, and Very Anglicized. This rating scale was a popular measure, often

used to assess acculturation and mental health status. In fact, the ARSMA was initially developed with the purpose of having a measure of acculturation that would serve as a moderator variable in clinical practice to help explain differences found in ethnic minority groups compared to the majority group (Cuéllar et al., 1995). The ARSMA measured four main factors: (a) the extent to which Mexican Americans used Spanish versus English along with their preference, (b) how they perceived their ethnic identity and to which classification they placed themselves, (c) cultural heritage and ethnic behaviors they exhibited, and (d) ethnic interaction (Cuéllar et al., 1980).

Cuéllar et al. (1995) noted several criticisms of the ARSMA, one major criticism being that acculturation was treated as unidimensional, or in a zero-sum manner, with Mexican culture and American culture on opposite ends of the linear representation. The assumption was that to move towards one cultural extreme was to lose aspects of the other culture. Additionally, the Likert scale provided no way of detecting marginalization—low identification with both cultures. Therefore, the ARSMA-II was developed to address these criticisms (Cuéllar et al., 1995). The ARSMA-II was a revised version that took a multidimensional approach to acculturation, based on a framework resembling Berry's (1980) four acculturation profiles rather than a linear representation of two cultural extremes.

With the development of these measures, however, there were still overarching issues not addressed by the measurement developers. Namely, authors incorrectly incorporated demographic indicators as part of the construct which thereby increased validity. Authors also tended to only use one Hispanic subgroup or validated some scales (e.g., ARSMA-II) using only college students, two practices which limit generalizability. Finally, with the inclusion of

multiple scales within a measure, data reduction techniques, such as factor analysis, were not used to decide which items belong to which scale (Marín & Gamba, 1996).

Marín and Gamba (1996) developed the Bidimensional Acculturation Scale for Hispanics (BAS) to address some of the methodological limitations in these previous scales. The primary components of the BAS that differentiate it from previous measures was that 1) participants were recruited and interviewed over the phone, thus not relying on college students, 2) a multidimensional framework was used in lieu of the linear conceptualization, and 3) subscales for different areas of acculturation were created using factor analysis, where previous measures either relied on single items for each domain, or did not use data reduction techniques to create scales within the measure (Marín & Gamba, 1996). The authors assert that the measure was valid for Mexican Americans and Central Americans.

Basilio et al. (2014) developed another biculturalism scale for Mexican American adolescents and adults from a psychological perspective. Basilio et al. (2014) made the argument that existing measures of acculturation were based on demographic variables, such as exposure to language, ethnic practices, and other cultural influences. To address these shortcomings, they proposed measuring biculturalism on psychological constructs (e.g., identity, behavior, beliefs, attitudes, values, worldview, and ability to respond to cultural cues). In particular, Basilio et al. (2014) designed a measure to assess the bicultural elements of emotion, behavior, and cognition. Someone who is bicultural is hypothesized to feel comfortable, behave appropriately, and perceive advantages to being associated with both cultural contexts. This framework also taps into cultural frame switching and offers a starting point in considering how to measure and conceptualize unconscious responses that can be triggered with cultural cues.

Asian Ethnic Groups

In 2001, the ratio of the number of measures for Hispanic groups to the number of measures for Asian Pacific American groups was 27:4 (Kim & Abreu, 2001). The most prominent measure for Asian American acculturation was the SL-ASIA (Suinn et al., 1987). Similar to the original ARSMA, the SL-ASIA assessed acculturation from a unidimensional standpoint. Given this limitation, Chung et al. (2004) developed the Asian American Multidimensional Acculturation Scale (AAMAS), which took a bidimensional approach to create an instrument that could be used with multiple Asian ethnic groups. In striving for applicability towards multiple Asian ethnic groups, Chung et al. (2004) raise the concept of pan-ethnicity, a third dimension not explored in previous studies.

Pan-ethnicity refers to a general “solidarity” among ethnic subgroups, often caused by categorization of subgroups to form one larger group (e.g., Koreans, Japanese, Thai, Malay, Indonesian, Taiwanese, Indian, etc., all collectively referred to as “Asians”). Pan-ethnicity is manifested when individuals of subgroups lay aside their subgroup differences to identify with the overall pan-ethnic culture (Chung et al., 2004). Pan-ethnicity is important to consider because individuals of subgroups, perhaps while acculturating to the majority culture, may also be acculturating towards a pan-ethnic culture. This means that they are not quite adhering to their culture of origin, rather, to a new culture only formed by a collection of subgroup interactions. Thus, the authors of the AAMAS strived to include a measure of pan-ethnic identification as part of acculturation. The measure consists of three scales: AAMAS-Culture of Origin (AAMAS-CO), AAMAS-Asian American (AAMAS-AA), and AAMAS-European American (AAMAS-EA).

The authors found that the AAMAS revealed a four-factor structure: cultural identity, language, cultural knowledge, and food consumption. A limitation of this measure was the

difficulty in validating pan-ethnic dimensions due to a lack of previous work on the concept. The authors hypothesized that pan-ethnicity would be related to two factors: (a) length of residence and opportunities to interact with other subgroups within the context of the host culture, and (b) derivation of some shared cultural base in similarity of experiences in the host culture. To demonstrate validity for the pan-ethnic dimension, the authors explained that the correlation patterns between the AAMAS-AA and the other previously validated measures showed similar patterns compared to AAMAS-CO, but to a lesser extent. For example, someone with a strong positive correlation between AAMAS-CO and SL-ASIA would also have a positive correlation between AAMAS-AA and SL-ASIA, but of a lesser magnitude (Chung et al., 2004).

Ethnic Minorities Around the World

Until this point, the majority of reviewed literature has existed in examining ethnic minority acculturation in the U.S. From a global standpoint, migration movements have led to increased cultural diversity in many countries. Relocating in any capacity increases the need for mental health practitioners to understand migration effects on psychological functioning, as well as how to interact and care for a more diverse population (Eytan et al., 2007). Additionally, Eytan et al. (2007) emphasized a need for an instrument that is, among other criteria, brief, multi-dimensional, bi-directional, and multiethnic. This measure was developed in Switzerland and validated with immigrant adults from Italy, Portugal, and Spain. One important aspect the authors brought forth was the potential to use this scale, not necessarily for people of the same ethnic groups, but for immigrants with similar motivations for migrating. This measure, in particular, was designed for immigrants who migrated based on economic reasons. It was also created for the more specific purpose of use in clinical purposes by mental health professionals.

Empirical

Acknowledging the two different frameworks for acculturation (unidimensional versus multidimensional), Abe-Kim et al. (2001) considered different scoring procedures as a way of considering unidimensional versus multidimensional perspectives. Asian American undergraduates were asked to take the SL-ASIA. The authors considered major cultural indicator variables (i.e., individualism-collectivism, independent-interdependent self-construal, loss of face, and impression management) that mark differentiation between the Asian and American cultures. Accordingly, the SL-ASIA was scored to either reflect the unidimensional or the bidimensional framework. The authors found that generational status can be used as a proxy only if acculturation is viewed from a unidimensional perspective. If, however, acculturation is considered the amalgamation of a set of cultural orientations, then generational status cannot and should not be used as a proxy variable.

In assessing acculturation using the bidimensional approach, researchers assume that identifying with one culture is not dependent on identification with the other culture. Therefore, there is independence, or orthogonality, in the maintenance of ethnic identity (Kang, 2006). There exist anomalies in the findings such that some studies revealed a violation of the independence assumption. Three scale formats are often used to assess different domains of acculturation. Frequency (e.g., “How much do you speak [language] at home?” “How often do you eat [culture] food?”) is used to assess a number of different domains, but is most often used to assess language use. Proficiency (e.g., “How well can you speak [language]?”) is also most often used to assess language proficiency. Endorsement (e.g., “I am proud of [culture] culture.” “I celebrate [culture] holidays.”) often measures attitudes, values, or preferences. Kang (2006) points out a few major findings of acculturation studies that may be affecting the lack of orthogonality in findings. First, frequency formats are most suited for measuring use. The issue,

however, is that acculturation is more than language use. Using proficiency formats often aggravates the problem, as proficiency in one language often triggers a lower proficiency in the other language. Previous research suggests, then, that the endorsement format would be the most appropriate way of validly measuring acculturation.

Kang (2006) tested the hypothesis that scale format affects orthogonality by giving a group of Asian American university students the General Ethnicity Questionnaire (GEQ; Tsai et al., 2000) to measure acculturation. The GEQ was selected because all three scale formats were included, which allowed for the hypothesis (endorsement vs. frequency/proficiency) to be tested. Additionally, the GEQ has two versions: GEQ-American (AM) and GEQ-Asian (AS), differing only in the reference culture. Examining correlations between the different sections and versions, the results revealed a small correlation between versions (AM vs. AS) for endorsement items, while frequency and proficiency formatted items had strong correlations between measure versions. They concluded that there is sufficient evidence to suggest frequency and proficiency scale formats are contributing to the lack of orthogonality found in some studies.

Previous research has shown the effects of priming using cultural icons and symbols on acculturation measures. Lechuga (2008) examined the effects of priming on bilingual Mexican American college students. Participants were recruited based on the ability to read, write, and speak both languages (English and Spanish). They were randomly assigned to receive either Mexican or American cultural primes (e.g., flag, icons, food, monuments, etc.) before being asked to fill out a questionnaire including measures of language proficiency, self-construal (collectivistic vs. individualistic views), and ethnic identity and acculturation/enculturation. The results support the hypothesis that priming affects how participants self-identify their ethnic categorization. The main limitation in this study, however, was its basis in the measure of

language proficiency. The authors misplaced emphasis on language proficiency as a measure of acculturation and failed to consider psychological constructs that may be affected by cultural priming.

The majority of research in acculturation has taken place in the U.S. or Canada with ethnic minority undergraduate students, but Chen et al. (2008) referenced the importance of considering different groups of people affected by migration trends. To understand the effects of contact with either the majority or minority ethnic culture on different groups, Chen et al. (2008) measured acculturation effects on psychological adjustment for three different groups: immigrants (Chinese immigrants permanently relocated to Hong Kong), sojourners (Filipino domestic workers in Hong Kong, temporarily relocated for an undetermined amount of time), and individuals of the majority culture who come into contact with a second cultural group or language within their original culture (Chinese and Hong Kong university students in multicultural multilingual settings). This study contributed knowledge for two groups rarely considered in previous literature: sojourners and individuals of the majority culture. Given that immigrants versus sojourners may have different motivations and perspectives on adjusting to the new culture due to differences in their length of stay, BII was an important factor to measure. The results showed that BII was a significant predictor of psychological adjustment for immigrants and individuals of the majority culture. The same effect of BII on psychological adjustment was not found for sojourners, potentially due to their indefinite length of stay. Instead, an important predictor of psychological adjustment for sojourners was performance-related skills, such as language proficiency.

Extending upon language proficiency and cognition frameworks, Schwartz et al. (2014) further examined cultural frame switching by randomly assigning bilingual first-generation (born

outside of the U.S.) and second-generation (parents born outside the U.S. but participant born in the U.S.) immigrant Hispanic college students to complete acculturation measures in either Spanish or English. The results showed measurement invariance across language of assessment. Latent mean differences, however, reveal slightly different results than what the authors initially hypothesized. For example, participants in the English language condition reported higher U.S. cultural behaviors as well as higher Hispanic cultural behavior than those in the Spanish language condition. Language use across the two domains (U.S. vs. Hispanic practices) did not differ across language conditions, which counters part of the hypotheses. Yet the authors remarked that despite evidence to suggest the appropriateness of pooling across language, these findings serve as a foundation for further research into specific conditions under which assessment language can be pooled. This conclusion was based on a fully bilingual sample within a highly bicultural context. Measurement equivalence found in this study may not be generalizable to individuals who are not fluent in both languages, do not live in a bicultural environment, or do not attend university (Schwartz et al., 2014).

Validity Issues

Participants are sometimes given the option to self-select into the language in which to complete the assessment. Some studies, however, ignore this self-selection factor and aggregate participant data, consequently also ignoring language (Schwartz et al., 2014). When studying acculturation, or a culturally based construct, there are four main threats to validity when self-selection into the language of the administered assessment is ignored (Schwartz et al., 2014). The four threats Schwartz et al. (2014) delineate are cultural frame switching, stereotype threat, translation quality, and language competency.

First, cultural frame switching occurs when people unconsciously switch their behavior, frame of thinking, language preference, etc. when primed with certain cultural cues. By priming people with certain contextual cues (e.g., language, symbols, behaviors, or mannerisms) from either their heritage culture or culture of exposure, their responses towards culturally related questions will differ. Answering questions given ambiguous stimuli is influenced by whichever cultural schemata has been activated. Benet-Martínez et al. (2002) demonstrated that BII was a moderator in the effect of priming cues, such that bicultural participants with high BII responded in a way that was congruent to both cultures. Conversely, there was a reverse priming effect for participants with low BII, who responded in a manner consistent with the opposite culture of the given priming cues. For example, participants with low BII who were given American cues responded with a characteristically Asian behavior.

Luna et al. (2008) conducted three studies that examined how language triggers frame switching in a Hispanic female sample. The researchers also considered implicit attitudes and conceptual associations between stimuli across languages, as there may be moderating effects of stimuli words across languages. They make the distinction between people who are bicultural versus people who are bilingual but not necessarily bicultural (monocultural). Bicultural refers to the internalization of two cultures, often implying proficiency and/or fluency in the associated languages (Luna et al., 2008). It is possible for individuals to learn a second language but are not subsequently considered bicultural. The results suggest that cultural frame switching only occurs for bicultural individuals; the same results were not found for bilingual monoculturals. This is because cultural frame switching involves, in addition to language, identity and behavior. Luna et al. (2008) argue that cultural frame switching is due to gained experiences which form a

cultural framework, a domain that cannot be built without direct experience of a language's cultural context.

Second, stereotype threat refers to anxiety or concern a person may experience that could potentially negatively affect their performance, thereby confirming a negative stereotype for that social group (Schwartz et al., 2014; Steele, 1997). Similar to cultural frame switching, stereotype threat involves activating schemata for a certain culture, which often includes the stereotypes associated with that group. In some cases, however, stereotype threat may promote avoiding both assimilation as well as upholding one's ethnic culture. Consequently, results affected by stereotype threat may be opposite to what cultural frame switching may predict.

Third, translation quality threatens validity in that different translations of a measure may not necessarily be valid. Back-translating is an oft used method for creating different translations of a measure. Although measures are taken to ensure the intended meanings of items are not altered, final scores taken across languages may not always have the same meaning. To further exasperate the problem, O'Bryon and Rogers (2010) report that self-identified bilingual school psychologists who administer language proficiency assessments do not clearly identify the level of fluency in the second language, and often, assessments administered in the second language are given with discomfort, indicating a lack of proper training and consideration of translation measures in acculturation assessments.

Fourth, language competency is a concern when self-selection is an option because there are always certain influences that affect selection of one assessment language over another. Some of these influences include, but are not limited to, language competency, comfortability, and social stigma. A likely reason for selecting a certain language is that the person is more fluent or comfortable in that language. On the other hand, there may be unspoken social

pressures to improve or demonstrate competency in a language, so the language is selected despite non-proficiency.

Discussion

A persisting trend in the literature is the difficulty in constructing a definitive way to measure acculturation. As there are different facets of culture, so there are also many ways of assessing acculturation. This is both a strength and a limitation. The strength lies in the continuous examination of acculturation from a myriad of perspectives, each researcher offering a slightly different view adds richness and diversity to a field so grounded in multiculturalism. With the many different subgroups to consider, there have been approaches for scale development. First, researchers can attempt to develop a different scale for every subculture. Given the vast number of subcultures, however, many minority cultures composed of a small number of members may be neglected. There is also the ever-present issue of generalizability to other similar subcultures. Second, the introduction of measuring acculturation towards a pan-ethnic identification seems to be a promising take on measuring individual subcultures. There is, however, always a risk of becoming too dependent on measuring pan-ethnic identity, attempts on overgeneralizing, and thus loss of levels in subcultural diversity. Third, changes in the direction of variable selection shifted from using demographic variables (e.g., language use, media preference, food selection, friends, etc.) to considering psychological constructs (e.g., emotions, cognition, attitudes, beliefs) to an integrated view of both (i.e., cultural frame switching). Understanding general psychological constructs and the cues that could be used for bicultural people to switch from one frame to another are conducive to gaining a more comprehensive picture of acculturation and/or biculturalism.

Implications

Although opportunities to interact with other cultures serve to enrich the human experience, there is the reality of disparities (e.g., economic, educational, health, etc.) between different ethnic groups that may be exasperated by dependency on testing for decision making if different cultural norms are not considered. Assessments are administered to make decisions for a variety of domains, including (but not limited to) assessing educational outcomes, aptitude, personality, mental and physical health, well-being, and job selection. Therefore, correctly assessing acculturation is a step in gaining a better understanding of how to harmonically interact in multicultural settings.

Limitations

As with any study, there remain gaps in the literature to be filled and limitations to be addressed. First, since measuring acculturation is often used to help determine needs for important aspects of people's lives, we cannot just focus on acculturation, per se. Rather, there is a need to understand that a bi- or multicultural identity exists. In order to best capture bi- or multicultural dimensions of identity, when dealing with people of different cultures, we must also account for language barriers, differences in proficiency of language, and comfortability in operating within certain cultural contexts, as cultural cues may trigger different frameworks for response (culture frame switching). Thereby, in order to capture the most valid information for making decisions, we have to be cognizant of the effects of language in assessment. One of the challenges, however, is that of bi- or multilingual people who exhibit different levels of proficiency in a language. Furthermore, there is the consideration of self-selection and reasons for opting to take an assessment using a certain language over another, which cannot be ignored (e.g., Schwartz et al., 2014).

Second, the literature lacks research studies that examine patterns of change, fluidity and transition between profiles. Of the reviewed studies, few, if any, incorporate statistical and methodological techniques that measure change in acculturation status *over time* (e.g., latent transition analysis, latent profile analysis). Along with using longitudinal designs, there is also a lack of consideration for external factors. One example is examining how societal stereotypes, attitudes, and perceptions of a certain culture shift over time (e.g., traditional aspects of a minority culture become heavily endorsed by the majority culture, consequently casting the minority culture in a positive light). Similarly, given the spread of the internet, the role and influence of digital media and social media (access to online resources, communities, cultural influences) should be examined. With digital media, representation of previously underrepresented ethnic minority groups may bring cultural awareness to the majority. With social media, people of ethnic minority groups have access to online resources, communities, and cultural influences that may serve to strengthen their ethnic group identification.

Third, acculturation, by definition, should occur for people of both the minority and majority ethnic groups through contact. Given current political disagreements in different multicultural contexts, it would be beneficial to consider how members of an ethnic minority group affect perceptions, attitudes, and opinions of members in the majority culture.

Unfortunately, not much work has been done on how cross-cultural interactions have affected members of the majority cultural group (with the exception of Chen et al., 2008). People of both the ethnic minority and majority make up the entire population; thus, changes in acculturation for the majority group when in contact with people from diverse cultures should be taken in to account to grasp a fuller understanding of interpersonal communication, adjustment, and social

processing. Understanding how mindsets shift with consistent exposure to an ethnic minority culture could serve to ameliorate strained political relations.

Overall, the articles reviewed in this paper suggest that trends and patterns in acculturation should be continually revisited. As globalization continues to influence human migration over time, so then must the measures for understanding acculturation be revised to reflect shifts in demographics, historical contexts, and cultural and societal perceptions and attitudes of different ethnic groups as people of different cultures continue to coexist, interact, and intermingle.

References

- Abe-Kim, J., Okazaki, S., & Goto, S. G. (2001). Unidimensional versus multidimensional approaches to the assessment of acculturation for Asian American populations. *Cultural Diversity and Ethnic Minority Psychology*, 7(3), 232-246. <https://doi.org/10.1037//1099-9809.7.3.232>
- Basilio, C. D., Knight, G. P., O'Donnell, M., Roosa, M. W., Gonzales, N. A., Umaña-Taylor, A. J., & Torres, M. (2014). The Mexican American Biculturalism Scale: Bicultural comfort, facility, and advantages for adolescents and adults. *Psychological Assessment*, 26(2), 539-554. <https://doi.org/10.1037/a0035951>
- Benet-Martínez, V., Leu, J., Lee, F., & Morris, M. W. (2002). Negotiating biculturalism: Cultural frame switching in biculturals with oppositional versus compatible cultural identities. *Journal of Cross-Cultural Psychology*, 33(5), 492-516. <https://doi.org/10.1177/0022022102033005005>
- Berry, J. W. (1980). Acculturation as varieties of adaptation. In A. Padilla (Ed.), *Acculturation: Theory, models, and some new findings* (pp. 9-25). Westview.
- Chen, S. X., Benet-Martínez, V., & Bond, M. H. (2008). Bicultural identity, bilingualism, and psychological adjustment in multicultural societies: Immigration-based and globalization-based acculturation. *Journal of Personality*, 76(4), 803-838. <https://doi.org/10.1111/j.1467-6494.2008.00505.x>
- Chia, A.-L., & Costigan, C. L. (2006). A person-centered approach to identifying acculturation groups among Chinese Canadians. *International Journal of Psychology*, 41(5), 397-412. <https://doi.org/10.1080/00207590500412227>
- Chung, R. H. G., Kim, B. S. K., & Abreu, J. M. (2004). Asian American multidimensional

- acculturation scale: Development, factor analysis, reliability, and validity. *Cultural Diversity and Ethnic Minority Psychology*, 10(1), 66-80. <https://doi.org/10.1037/1099-9809.10.1.66>
- Comas-Diaz, L., & Grenier, J. R. (2002). Migration and acculturation. In J. Sandoval, C. L. Frisby, K. F. Geisinger, J. Scheuneman, & J. R. Grenier (1997), *Test interpretation and diversity: Achieving equity in assessment* (pp. 213-240). American Psychological Association.
- Cuéllar, I., Arnold, B., & Maldonado, R. (1995). Acculturation rating scale for Mexican Americans-II: A revision of the original ARSMA scale. *Hispanic Journal of Behavioral Sciences*, 17(3), 275-304. <https://doi.org/10.1177/07399863950173001>
- Cuéllar, I., Harris, L. C., & Jasso, R. (1980). An acculturation scale for Mexican American normal and clinical populations. *Hispanic Journal of Behavioral Sciences*, 2, 199-217.
- Cummins, J. (1982). Tests, achievement, and bilingual students. *National Clearinghouse for Bilingual Education*, 9, 2-9.
- Diana v. California State Board of Education, No. C-70, RFT, (N. D. Cal. 1970).
- Eytan, A., Jene-Petschen, N., & Gex-Fabry, M. (2007). Bicultural identity among economical migrants from three south European countries living in Switzerland. Adaptation and validation of a new psychometric instrument. *BMC Psychiatry*, 7(17). <https://doi.org/10.1186/1471-244X-7-17>
- Figueroa, R. A. (1989). Psychological testing of linguistic-minority students: Knowledge gaps and regulations. *Exceptional Children*, 56(2), 145-152. <https://doi.org/10.1177/001440298905600206>
- Frisby, C. L. (1998). Culture and cultural differences. In J. Sandoval, C. L. Frisby,

- K. F. Geisinger, J. Scheuneman, & J. R. Grenier (Eds), Test interpretation and diversity: Achieving equity in assessment (pp. 51-73). American Psychological Association.
- Kang, S. M. (2006). Measurement of acculturation, scale formats, and language competence: Their implications for adjustment. *Journal of Cross-Cultural Psychology*, 37(6), 669-693.
<https://doi.org/10.1177/0022022106292077>
- Kim, B. S. K., & Abreu, J. M. (2001). Acculturation measurement: Theory, current instruments, and future directions. In J. G. Ponterotto, J. M. Casas, L. A. Suzuki, & C. M. Alexander (Eds.), *Handbook of multicultural counseling* (2nd ed., pp. 394-424). Sage.
- King, J., & Keane, E. (1992). *Acculturation Scale for American Indian Adolescents: Voices of Indian teens project*. University of Colorado Health Sciences Center.
- Kroeber, A. L., & Kluckhohn, C. (1952). *Culture: A critical review of concepts and definitions*. Vintage Books.
- Lau v. Nichols, 414 U.S. 563 (1974).
- Lechuga, J. (2008). Is acculturation a dynamic construct? The influence of method of priming culture on acculturation. *Hispanic Journal of Behavioral Sciences*, 30(3), 324-339.
<https://doi.org/10.1177/0739986308319570>
- Luna, D., Ringberg, T., & Peracchio, L. A. (2008). One individual, two identities: Frame switching among biculturals. *Journal of Consumer Research*, Inc., 35, 279-293.
<https://doi.org/10.1086/586914>
- Marín, G. & Gamba, R. J. (1996). A new measurement of acculturation for Hispanics: The Bidimensional Acculturation Scale for Hispanics (BAS). *Hispanic Journal of Behavioral Sciences*, 18(3), 297-316. <https://doi.org/10.1177/07399863960183002>
- McLean, J. E. (1995). *Improving education through action research: A guide for administrators*

and teachers. Corwin Press.

Moyerman, D. R., & Forman, B. D. (1992). Acculturation and adjustment: A meta-analytic study. *Hispanic Journal of Behavioral Sciences*, 14(2), 163-200.

<https://doi.org/10.1177/07399863920142001>

O'Bryon, E., & Rogers, M. R. (2010). Bilingual school psychologists' assessment practices with English language learners. *Psychology in the Schools*, 47(10), 1018-1034.

<https://doi.org/10.1002/pits.20521>

Padilla, A. M. (2001). Issues in culturally appropriate assessment. In L. A. Suzuki, J. G. Ponterotto, & P. J. Meller (Eds.), *Handbook of multicultural assessment* (2nd ed., pp. 5-27). Jossey-Bass.

Rohner, R. P. (1984). Toward a conception of culture for cross-cultural psychology. *Journal of Cross-Cultural Psychology*, 15, 111-138. <https://doi.org/10.1177/0022002184015002002>

Sandoval, J. (1998). Testing in a changing world: An introduction. In J. Sandoval, C. L. Frisby, K. F. Geisinger, J. Scheuneman, & J. R. Grenier (Eds.), *Test interpretation and diversity: Achieving equity in assessment* (pp. 3-16). American Psychological Association.

<https://doi.org/10.1037/10279-015>

Schwartz, S. J., Benet-Martínez, V., Knight, G. P., Unger, J. B., Zamboanga, B. L., Des Rosiers, S. E., Stephens, D. P., Huang, S., & Szapocznik, J. (2014). Effects of language of assessment on the measurement of acculturation: Measurement equivalence and cultural frame switching. *Psychological Assessment*, 26(1), 100-114.

<https://doi.org/10.1037/a0034717>

Schwartz, S. J., & Unger, J. B. (2010). Biculturalism and context: What is biculturalism, and when is it adaptive? *Human Development*, 53, 26-32. <https://doi.org/10.1159/000268137>

- Schwartz, S. J., & Zamboanga, B. L. (2008). Testing Berry's model of acculturation: A confirmatory latent class approach. *Cultural Diversity and Ethnic Minority Psychology*, 14(4), 275-285. <https://doi.org/10.1037/a0012818>
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52, 613-629.
<https://doi.org/https://psycnet.apa.org/doiLanding?doi=10.1037/0003-066X.52.6.613>
- Sue, D. W., & Sue, D. (1990). *Counseling the culturally different*. Wiley.
- Suinn, R. M., Richard-Figueroa, K., Lew, S., & Vigil, P. (1987). The Suinn-Lew Asian Self-Identity Acculturation Scale: An initial report. *Educational and Psychological Measurement*, 47, 401-407. <https://doi.org/10.1177/0013164487472012>
- Szapocznik, J., Scopetta, M. A., Arnalde, M., & Kurtines W. (1978). Cuban value structure: Treatment implications. *Journal of Consulting and Clinical Psychology*, 46, 961-970.
<https://doi.org/10.1037/0022-006X.46.5.961>
- Triandis, H. C., & Brislin, R. W. (1984). Cross-cultural psychology. *American Psychologist*, 39, 1006-1016. <https://doi.org/10.1037/0003-066X.39.9.1006>
- Tsai, J. L., Ying, Y. W., & Lee, P. A. (2000). The meaning of "being Chinese" and "being American": Variation among Chinese American young adults. *Journal of Cross-Cultural Psychology*, 31, 302-332. <http://dx.doi.org/10.1177/0022022100031003002>
- U.S. Census Bureau. (2018). QuickFacts. Retrieved from
<https://www.census.gov/quickfacts/fact/table/US/PST045218> .
- Van de Vijver, F. J. R., & Phaet, K. (2004). Assessment in multicultural groups: The role of acculturation. *Applied Psychology: An International Review*, 53(2), 215-236.
<https://doi.org/10.1111/j.1464-0597.2004.00169.x>

The Current State of Assessing Historical Thinking: A Literature Analysis

Taylor Hamblin

Department of Teaching, Learning, and Teacher Education
University of Nebraska-Lincoln

Abstract

In American schools since the mid 2000's, social studies departments and state departments of education have created goals and updated standards prioritizing critical thinking engagement. Promotion of critical thinking has created a wealth of scholarship on developing a specific type of critical thinking, or cognition, called historical thinking. Imperative to the promotion of teaching historical thinking is in how teachers can assess the inquiries that make it up. Unfortunately, standardized social studies assessments have failed to measure the acquisition of the new historical thinking standards. In order to improve the assessment practices of history teachers, I wish to do two things: (1) switch the focus from recall-memorization assessments to those that will focus on a diverse array of historical thinking inquiries; and (2) improve the assessments that we currently use to measure historical thinking skills. In order to accomplish these two objectives, I will examine empirical research studies that focused on how students obtain historical thinking skills. From the data, I ascertain how practicing teachers and researchers currently measure historical thinking skills, and from that, propose improvements.

Keywords: historical thinking, social studies education, history education, assessment, cognition

doi: 10.32873/unl.dc.ne008

In American schools since the mid 2000's, social studies departments and state departments of education have created goals and updated standards prioritizing critical thinking engagement. Promoting critical thinking is not unique to social studies, but it has created a wealth of scholarship on developing a specific type of critical thinking called historical thinking or historical reasoning. This type of thinking involves using inquiry to answer historical questions, like "how important was religion in Colonial America?", and evaluate social concepts, like religion, global citizenship, or cause & effect. In addition, many social studies educators want historical inquiry to focus on "real world" questions that students truly want to know (Selywn, 2014). Unfortunately, standardized social studies assessments have failed to measure the acquisition of the new standards and goals because teachers (or anyone who assesses) historical thinking skills are unlikely to be successful with current practices and assessment tools (Reich, 2009; Shemilt, 2018).

Difficulties arise from the cost of assessing higher order thinking, preparing preservice educators to competently teach social studies cognition, and changing traditional testing practices (Shemilt, 2018). Despite the struggle of assessing cognitive processes in social studies courses, there is tremendous value in creating these assessments. Improved cognitive processes in social studies courses could result in improved civic engagement, student motivation & agency, and better career prospects for students (McGrew, 2018, National Council for the Social Studies, 2013, Selwyn, 2014). An example of assessment that promotes civic engagement is when my 8th grade students created bills for a Mock Senate. During the Spring of 2019, my teenage students questioned, researched, debated, and voted on bills that they created: these bills went onto school admin for consideration. While students completed these bills, I assessed them to determine their

ability to source documents, corroborate evidence, and make inferences over arguments and reasoning (Hamblin, 2019).

Despite my success of assessing students' cognitive abilities while they completed a project, the previously mentioned issues hinder the assessment of historical thinking. To examine this issue, I searched for scholarly empirical studies that examined historical thinking assessments or examined how historical thinking could be measured. In addition, I focused on scholars who were research leaders in studying historical thinking, such as Sam Wineburg, Peter Seixas, John Lee, Bruce VanSledright, David Hicks, and Peter Doolittle. From these searches I found studies defining historical thinking, or reviewing the philosophy of it, but I found none that directly examined how assessments could be scaffolded for students. A lack of finding research examining scaffolding is troubling because any assessment of cognitive ability must follow some type of growth model, like Jerome Bruner's *spiraling curriculum* (Bruner, 1960).

According to the glossary of education reform, "scaffolding refers to a variety of instructional techniques used to move students progressively toward stronger understanding and, ultimately, greater independence in the learning process" (2015). Scaffolding is a means to ease students into a concept, curriculum, behavior, or idea and is a fundamental part of teaching and assessing historical thinking. As an example, let's say an elementary school teacher wants to teach their second graders how to find helpful sources to answer questions. This teacher might have a specific current event that they want to cover, but they will first break-up the lesson into interactive activities that teach the young students important concepts, like finding good sources. A teacher could ask students to write down the five most trustworthy people that could describe the student's identity and experiences. Students would enjoy writing down family members, teachers, and close friends. Finally, the teacher asks why the people listed are trustworthy

sources. This interactive activity then expands until students complete the standard, such as the C3 Framework's D1.5.K-2 "Determine the kinds of sources that will be helpful in answering compelling and supporting questions" (National Council for the Social Studies, 2013, p. 25).

Based on my analysis of research that examined assessment of historical thinking, there are five questions that are important to answer: (1) How do we currently measure historical thinking? (2) How can we take valid data from assessments over inquiry and historical thinking? (3) What scaffolding exists when testing for historical thinking and how can we improve it? (4) How does inquiry and assessment motivate students? (5) Why is assessing historical thinking and inquiry-based learning difficult? Through this literary analysis, I will answer these five questions and will use one to two empirical research article(s) to serve as an exemplar for answering each question.

Before answering the questions, it is foremost to provide context of the current state of historical thinking in academic research. Over the last twenty years, researchers have developed assessments for historical thinking. Assessing historical thinking is part of a broader movement of inquiry assessments. Inquiry assessment is not isolated to just history and the social studies disciplines; Stephanie Corliss (2011) examined how her science projects, which required students to solve basic to complex problems using a rational system, improved student content knowledge and their ability to apply concepts over multiple disciplines. If multiple disciplines are moving towards inquiry assessment, then there must be a broader force pushing social studies educators to redefine their purpose in education.

Fortunately, there are institutions and individuals researching historical thinking skills and producing teaching resources based on their findings. The founders and leaders of these groups have published several theoretical and practical studies that help define historical thinking

and the measurement of it. For example, Peter Seixas, a leader in studying cognitive abilities in history, wrote “A Model of Historical Thinking” (Seixas, 2012) which defined and explained historical consciousness, historical thinking, and the teaching of history in Canadian and United States schools. Seixas and others have defined historical thinking and created teaching resources, but the definitions are heterogeneous, and sometimes contradictory. Furthermore, the analysis they provide in their studies does not consider how historical inquiry assessments need to be scaffolded. An absence of analysis into inquiry assessment and scaffolding challenges the likelihood that teachers will use historical thinking as a curriculum and pedagogical tool.

The move to promoting more inquiry-based assessments, like testing for historical thinking, became clear because of the creation of the C3 Framework (2013). The C3 Framework is a set of standards that unites the disciplines of History, Civics, Geography, and Economics. Assessment designers created the C3 Framework to help improve students’ abilities to “recognize social problems; ask good questions and develop robust investigations into them; considering possible solutions and consequences; separate evidence-based claims from parochial opinions; and communicate and act upon what they learn” (National Council for the Social Studies, 2013, p. 6).

The objectives within C3 provide teachers with a structure to vertically align social studies disciplines across multiple grades. Missing from the framework is assessment advice, measuring devices, and scaffolding. The authors of C3 admitted this gap in their framework, and believed it would be “smart, thoughtful, and imaginative teachers” who need to find ways to make the framework adaptable (NCSS, 2013, p. 15). It is wise for education policy and standard makers to collaborate with teachers to make decisions on implementing curriculum in the

classroom. However, for this to work to be successful, school districts must give teachers time during the workday to be lucrative in producing implementation strategies.

Unfortunately, the C3 Framework lacks practical advice for teachers who may feel overburdened with other educational duties. John Lee was one of the writers for C3 and has developed inquiry assessments to make up for the gap. In an unpublished chapter titled “Assessing Inquiry,” Lee created a lesson plan, which focused on inquiry assessment, compelling questions, supporting questions, formative questions, featured sources, summative performance tasks, and taking informed action (Lee, manuscript submitted for publication). I include Lee’s unpublished chapter because it demonstrates that creating inquiry assessment is an ongoing issue. Lee’s lesson plan is a step in the right direction as it gives practicing teachers plenty of practical advice. Lacking is scaffolding advice.

The writers of the C3 Framework, and later Lee in his unpublished article, create lesson plans for general education students who are supposedly at the same academic level. This lack of differentiation makes it difficult to implement these lessons because of the realities of teaching. In order to be successful in my own practice, I felt that I needed to differentiate within the classroom environment through student choice, alternative primary sources, second- and third attempts on summative assessments, and choice boards. Going into the future, researchers need to do more to advise teachers and give resources on scaffolding historical thinking assessments.

Question 1: How do we currently measure historical thinking skills?

First, there is still confusion in what historical thinking skills are. In order to measure historical thinking skills, there must be categorizations of the different types of cognitive processes which comprise it. A clear breakdown is especially important, because as Stephane Levesque and Penny Clark pointed out in their handbook chapter about historical thinking

definitions; "...if the ability to think historically should go beyond the mere mastery of factual knowledge about the past ("know that"), it is still unclear as to what the alleged connections between "history" and "thinking" actually means in conceptual and practical terms ("know how") (2018, p. 119). Fortunately, researchers have been clarifying the connection between history and thinking by classifying historical thinking skills. For example, Peter Doolittle and David Hicks breakdown these skills into six separate inquiry categories (Virginia Tech), which are summarizing, contextualizing, inferring, monitoring, corroborating, and interpretation. Doolittle and Hicks are not the only researchers to promote a set of skills, Wineburg also did, and included skills that are different, like sourcing, which focuses analysis on a document or artifact's creator and the circumstances of its creation (Wineburg, 2001).

Although we have researchers describing the "know how" of historical thinking, and corresponding skills, there does need to be a greater effort in separating those skills so teachers may test for them. If researchers do not further bracket these skills, then it will be as Denis Shemilt points out, "it may not be possible to make secure assessments of students' historical consciousness," which impacts one's ability to historically think (2018, p. 453). For instance, Sam Wineburg and Sara McGrew, the author of the upcoming exemplary article in this paper, created activities that test specific skills (McGrew et al., 2018). One such assessment measures a student's ability to source a painting (Stanford History Education Group). The assessment only requires students to successfully identify that the painter created the painting at a different date than the event that the painting depicts (in this case the First Thanksgiving). Sourcing, and other historical thinking skills are more complicated than simply identifying a date; therefore, there needs to be further categorization of each skill (Appendix A).

Although there are descriptions of inquiry types and skills for historical thinking most history assessments do not adequately measure them, because they primarily utilize multiple choice questions to test factual recall. The questions on these standardized tests, at their best, only measure aspects of factual recall, but often even fail at doing that (Reich, 2009). The popularity of multiple choice is not surprising since they are simple to use, and as Denis Shemilt points out in his article *Assessment of Learning in History Education* (2018), teachers believed multiple choice tests “improved the reliability” of assessments (p. 449). There is potential for multiple choice tests to measure historical thinking, Bruce VanSledright argued convincingly that they could in his book *The Challenge of Rethinking History Education* (2011), but the issues in creating a reliable multiple-choice assessment are the same as creating any assessment that examines historical thinking. Creating, administering, and grading appropriate exams is expensive, as well as being time consuming. Current practicing teachers are unaware or unprepared to create and administer such assessments. In addition, researchers like Gabriel Reich suggest that recall tests cannot accurately measure historical knowledge (Reich, 2009). When primarily using multiple choice assessments, teachers are like a baker measuring ingredients using inches and feet, they are using the wrong measurement to assess historical thinking.

A popular alternative to factual recall tests is document-based questions (DBQs), which prompt students to analyze several primary documents, form a thesis, and defend it. In social studies courses, the most widespread use of DBQs is in the U.S. History Advanced Placement exam. If multiple choice recall tests are like measuring how much flour goes into the bowl in inches, then DBQs are using a jackhammer to mix the ingredients. DBQs are substantial questions, and require over an hour to complete; therefore, students use several types of inquiry and skills to form their answers. Due to the extent of these questions, it is unclear as to what

particular historical thinking skill is being measured. Analysis of questions that test historical thinking and what they actually measure is necessary because assessments have a tremendous influence over classroom curriculum and pedagogy. Since assessments are necessary, some researchers examined how they can be created in order to properly measure historical thinking.

Exemplar Article #1

Sarah McGrew and her team of Stanford researchers created short assessment tasks that measured students' ability to search for, evaluate, and verify online information. McGrew et al. work within the Stanford History Education Group (SHEG), which has developed online assessment tools that measure students' historical thinking ability. McGrew created assessment tasks for her research study "Can Students Evaluate Online Sources? Learning from Assessments of Civic Online Reasoning" (2018), which were similar to history lesson plans that SHEG authors created.¹ Although the study focuses on civic reasoning, many of the inquiries are similar to historical thinking because they use many of the same inquiry skills found within the historical thinking skills, like sourcing a document or artifact.

McGrew created fifteen assessment tasks, which focused on three constructs; "Who is behind the information? What is the evidence? What do other sources say?" These questions are similar to Wineburg's (2001) historical thinking skills, which are respectively, sourcing, critical thinking, and corroboration. 405 middle school students, 348 high school students, and 141 college students from twelve different states comprised the participant pool, and researchers collected 2,616 responses from this group. Once given the analysis sheets and online sources, students struggled to successfully evaluate online claims, sources and evidence. McGrew believed curriculum materials must be better in order to support students "civic online reasoning

¹ These assessment tasks can be seen at <https://sheg.stanford.edu/history-assessments>.

competencies” (McGrew, 2018, p. 165-166). In the case of the McGrew study, “better” seemed to mean “just need to exist.”

The McGrew study is helpful in understanding how to measure historical thinking skills because it provides a measuring framework for assessing online civic skills. McGrew measured civic skills in a similar way to how SHEG researchers did historical thinking skills. For example, McGrew used the exact same categories, such as “sourcing” that SHEG used in their Beyond the Bubble history assessments. It is beneficial to use the same types of skills in both civics and history because it means educators can use common assessments across social studies disciplines. Having common assessment measurements benefits curriculum like the C3 Framework because it demonstrates how it is possible to vertically align assessment across social studies disciplines.

Along with the benefit of vertical alignment, assessments that measure skills are necessary because students misinterpret information. McGrew identified commonly held misinterpretations of students, such as always trusting “news” sources, even when they are clearly biased (2018, 193). These types of misinterpretations exist across social studies disciplines. If teachers want to ensure their assessments are going to improve cognitive processes, then they must collaborate in how they teach and assess skills. This collaboration is imperative as students only have an average of one year of civics and will take history sporadically throughout their secondary education. SHEG and McGrew are developing reasoning and thinking skills, which could be the uniting force to vertically align social studies departments. Researchers must continue to improve and model pedagogical methods like scaffolding to encourage historical thinking as a standard practice.

Exemplar Article #2

Gabriel A. Reich has a high school history teacher background, where he grew increasingly frustrated with standardized testing over historical knowledge. Reich is also a historian, who focuses on how Americans, especially the young, learn about the Civil War and how myths form a significant part of our historical consciousness. Due to his background and frustrations, it is not a surprise that Reich examined how high school students choose answers on a set of multiple-choice questions. Reich wanted to know if students used historical reasoning (Reich's term) when they selected A, B, C, or D. The researcher focused on a class of urban 10th grade students who had to take a high-stakes exam at the end of the year in order to earn a social studies credit that they needed to pass high school. Reich used the questions from New York State's Global History and Geography Regents Exam, which is a required test in order to complete high school.

Based on students' answers, and interviews with students afterward, Reich determined that students were using test-wise thinking skills to select correct answers. For example, in many of the interviews, the researchers found that students would eliminate answers because they used a similar response on a different question, or the student knew a certain name did not fit the era they were studying. In these situations of answering test questions, historical thinking and factual recall played little part in how students choose correct answers. Reich did not see students using skills like sourcing, corroboration, continuity, or contextualization when they answered their multiple-choice questions.

Reich's research suggests that multiple choice tests do not accurately measure students' knowledge of history. It would be useful to use Reich's methods of examining multiple choice questions to examine the tests that VanSledright (2011) created. VanSledright created multiple

choice answers that he weighted, some were more correct than others, and students analyzed all of the answers to choose the most accurate one. Also, Reich developed useful categorizations for the skills that students used to select the correct answer. The skills were “test-wiseness,” “literacy,” and “domains-history content.” If the question is “how do we measure historical thinking skills,” then it is beneficial to know what may interfere with measuring them and what does not qualify as a skill. If students use skills outside of historical thinking to answer questions, then researchers and teachers must identify and manage those skills in a way where they do not interfere with assessment.

Reich’s study serves as a warning sign before designing standardized multiple-choice exams. The research suggests that multiple choice tests do not assess historical thinking, but they also do not accurately measure a student’s ability to recall content information. The ineffectiveness of standardized multiple-choice assessments is not a new revelation, but it is important to state because governments and teachers subjugate millions of students to these exams each year. This is especially disturbing in social studies courses because many states do not require standardized assessments of historical thinking or knowledge of history, but many social studies teachers still rely on exams that solely contain simple recall of factual information. Despite not having standardized tests, the testing culture has adversely affected social studies teachers enough where they have created their own standardized testing regime.

Question 2: How can we take valid data from assessments over inquiry and historical thinking?

In the last section analyzing the first question, I mentioned how DBQ assessments cannot accurately measure specific cognitive processes, like the historical thinking skill “sourcing.” Additionally, Reich demonstrated how students use unintended thinking processes to answer multiple choice questions. Exacerbating the issue of creating tests, confounding factors can

compromise the data pulled from tests. For example, Adam Wallace examined motivation and belief in oneself when examining National History Day projects (Wallace, 1987), in which Wallace suggested that students who believed in their historical thinking ability did better on assessments. Additionally, Lee and Ashby focused on students' interpretation of tone, theme, and timescale as they grew older; therefore, a student's age might compromise their ability to think historically. Due to the complexity of assessing historical thinking, assessment creators must be deliberate in identifying what they are measuring and the possible issues with their evaluations.

Exemplar Article #1

Sam Wineburg has been working on being deliberate since the late 1990s; he has routinely published work on the topic over the last 30 years (Wineburg, 1991; 1997, 2001; 2009; Smith et al., 2019). Wineburg received his doctorate in Psychological Studies in Education, which is in part why he is thoroughly invested in investigating how students think about and learn history. Examples of Wineburg's investment are his numerous publications, one of which is his book *Historical Thinking and Other Unnatural Acts: Charting the Future of Teaching the Past* (2001). In the book, Wineburg examines many reasons why students think about the past differently, such as the power of one's gender on historical thinking, which may affect how students measure when completing assessments that test historical thinking.

Wineburg wrote a chapter called "Picturing the Past" (2001) where he focused on the question: "how do boys and girls picture the past?" The researcher asked students to draw pictures of different historical figures, like Pilgrims, Western Settlers, and Hippies, in order to see how they pictured these people. Wineburg quantified the images based on gender, number of people, and types of actions the historical figures were committing. Additionally, Wineburg and his assistants conducted interviews to let students explain their reasoning behind the drawings.

Male students drew predominantly male characters, isolated or alone, and were more likely to be engaged in violence. Girls were more likely to draw female characters than boys, but their female to male ratio was 50/50. Girls also drew more groups of people such as families. Most concerning was the girls' propensity to fill their “historical world” with more men than would be realistic, like drawing a community square with only men. The researchers wondered, do girls do this when they are reading textbook accounts of historical events. If girls and boys have different outlooks of gender in history, how might this affect their historical thinking? Would it be fair to examine boys and girls using the same prompts and rubrics?

If young girls and boys are “seeing” a different historical world, then they will likely interpret primary sources and historical arguments differently. If this is the case, then gender may benefit or hinder students’ mastery over concepts. For example, a female student may be less likely to disassociate violence or discrimination from other actions that historical people committed. This is problematic because girls may conclude historical figures and events are invalid sources to use in arguments because they do not meet present moral standards, like American Founding Fathers owning slaves, Free Blacks settling on Native American lands, blaming Adolf Hitler for Germany’s anti-Semitism. Girls are not the only ones at risk, students from certain religious groups could believe that historical figures such as Thomas Jefferson or Martin Luther King Jr., who despite their many accomplishments, most likely committed adultery.

When designing assessments, creators must be careful to isolate cognitive processes from each other. Test creators can bracket historical thinking skills by breaking down current historical thinking inquiry types, like sourcing, into several different sub-skills (Appendix A). One such sub-skill is “identified the category of the source.” Assessments must explicitly

measure this sourcing subskill, along with the six others, in order to get a more accurate understanding of how well students are doing with the more general skill of sourcing. Furthermore, there needs to be an understanding of how students stumble into theoretical pitfalls, and how mastery may look differently for various groups of students. Due to these reasons, researchers should design and research scaffolds in order to help students reach mastery over skills and concepts. For example, a simple pedagogical scaffold would be to model analysis of primary documents.² A curriculum scaffold would be to have multiple types of the same reading, which would differ in reading level and/or theme.

Question 3: What scaffolding exists when testing for historical thinking and how can we improve it?

Scaffolds must be incorporated when researching the mastery of historical thinking skills. Scaffolds are the curriculum designs or pedagogical methods used to help students reach mastery over a certain skill or set of information and producing scaffolds is a time-consuming challenge. Creating scaffolds does not stop at teachers adapting curriculum for differences in reading levels or learning disabilities. There are additional socioeconomic, political, and natural circumstances affecting students that teachers must address by scaffolding curriculum. A United States teacher may need to adapt curriculum for students who do not speak English, are from cultural groups which represent “the enemy” in dominant, conservative, American narratives, like American Indians and Muslims, and students who simply cannot afford to ride the bus in the winter (Attewell, 2011, Renn, 2013). There are additional circumstances than the ones listed and teachers knowing about these confounding factors does not help when they have little time to build appropriate scaffolds. The education system is unlikely to change soon enough to meet the

² An example of this modeling can be watched at <https://www.youtube.com/watch?v=ib8R6T4qsJM&t=54s>

needs of diverse learners; social studies educators must individually implement curriculum and pedagogy that will help scaffold assessment measuring efforts.

After reviewing fifteen research studies over creating inquiry assessments, I found that some researchers address scaffolding in some way. For instance, in Monika Waldis's (2015) research study, German students mastered thinking skills with greater ease when they analyzed history that they were familiar with, such as "The Nazi Boycott of Jewish Businesses," compared to "Trade Relations with Japan." Waldis found students are more comfortable using historical thinking skills with familiar topics, which can support their willingness to take theoretical leaps necessary for historical thinking. Waldis concluded that teachers should use familiar topics as a scaffold to help students learn new historical thinking skills. Although Waldis and a few others mention scaffolding assessments of historical thinking directly or indirectly, researchers have not considered scaffolding enough to help practitioners create learning models for differentiated classrooms.

The lack of consideration towards practical application of scaffolding historical thinking assessments underlies a deeper issue. Teachers are struggling to apply historical thinking to their curriculum because their students are at different cognitive levels and possess different identities and backgrounds. For example, teachers may have a difficult time applying Waldis's research because there was no suggestion as to how assessments could be familiar to all students. Waldis did not describe how her finding of *familiarity promoting superior cognitive development* could apply to helping students progress through a historical thinking model. The two following exemplar articles also demonstrate how researchers did not consider the differentiated needs of students nor ideas on scaffolding while researching historical thinking.

Exemplar Article #1

Examination of the article “Fostering Analysis in Historical Inquiry Through Multimedia Embedded Scaffolding” (2008) suggests researchers do not consider how familiar students are with the historical topic. In their article, David Hicks and Peter Doolittle developed a strategy for historical thinking called SCIM-C. The strategy stands for “summarizing,” “contextualizing,” “inferring,” “monitoring,” “corroborating,” and “interpretation.” These historical thinking skills are similar to how Sam Wineburg, Sarah McGrew, and Monika Waldis define their skills. Students and teachers can use the SCIM-C strategy as a rational system to answer historical questions and analyze documents and artifacts. First publishing their ideas for the strategy in 2004, Hicks and Doolittle join other scholars in arguing for a greater evidence-based approach in history education. In their study, Hicks and Doolittle report their findings and answer the question; does the SCIM Historical Inquiry Tutorial foster the development of historical source analysis?

Seventy-seven college undergraduates compromised the study; they were enrolled in a general studies health education course. Researchers chose participating students from the health course because they would have little knowledge of historical procedures. The study introduced the SCIM strategy (the researchers removed the C for this study) to the students over three instructional periods, and researchers assessed student’s knowledge using a *single* open-ended question. Based on the teaching of the SCIM strategy, many students applied their newfound skills as part of a cognitively sophisticated process of analyzing sources.

Despite the success of numerous students, Hicks and Doolittle found that students applied historical thinking skills unevenly. This unevenness could be due to students not receiving differentiated assessments. Using Monika Waldis’s theory on familiarity, students could have

underperformed because they lacked knowledge of the historical time period used for the assessment questions. There is a chance that other confounding factors could have skewed data, like the factors mentioned under the second question of this paper. Whether it was lack of familiarity with historical topics, or something else, researchers and teachers should consider scaffolds to help students equitably reach mastery. If this is the case, then Hicks and Doolittle have data that does not truly show mastery of isolated thinking skills, but more a relationship between skills and knowledge of historical content. If K-12 teachers used the same strategy as Hicks and Doolittle to assess younger students, then they would need to differentiate the content in the assessments. Just one of these differentiated scaffolds would be allowing students to choose content that the teacher will use to assess them.

Exemplar Article #2

Along with familiarity with a topic, the identity and background of someone can affect their ability to master historical thinking skills. Instead of seeing knowledge and perspective as affecting the ability to historically think, Peter Lee and Rosalyn Ashby attributed age as a more prominent factor in their study “Progression in Historical Understanding Among Student Ages 7-14” (2000). As one of their central tasks, Lee and Ashby examine how students change their perceptions of history as they age. The philosophy of this research falls in line with Jean Piaget’s ideas of students learning through a cognition model as they grow older, with strict limits on what a student can do at a certain age.

In the main investigation, Lee and Ashby collected responses from 320 children between the ages of seven and fourteen. They also interviewed 1/3 of the students in order to determine the reasoning behind their interpretations of history. Students responded to questions by examining secondary source accounts of Romans in Briton, but each story differed in theme,

tone, and timescale. As students got older, they described differences in the stories based on their dates and abstract concepts. Lee and Ashby took this observation and theorized that students' progress in their formation of history as they age. Lee and Ashby use several practical codes when measuring students' historical thinking. Some of these codes were "selection," "legitimate viewpoint," "intentional distortion," "mistakes," and "opinion unexplained" (Lee & Ashby, p. 58).

The disadvantage with Lee and Ashby's findings is they do not consider how confounding factors like familiarity with a topic, student perspectives, or the amount of knowledge a student possesses may affect their ability to historically think. Simply using age as a factor in how students develop prevents educators and researchers from developing scaffolds to assist students in mastering historical thinking skills. Since Lee and Ashby believe age is a factor in how well students can understand abstract concepts, they perhaps did not see a reason to formulate any steps or methods that would help students progress through a cognitive learning model. They also did not apply existing models, like Jerome Bruner's *spiralizing curriculum*, which has been effective in helping students learn deeper concepts (Bruner, 1960). Researchers benefit from a model like Bruner's because if a researcher or teacher is able to measure progression of historical thinking, then scaffolds can be likely built between each step to help with the advancement of cognition.

Question 4: How does inquiry assessment motivate students?

A reason teachers scaffold a lesson or activity is because it motivates students to accomplish tasks which lead to mastery over skills, concepts, and information. In essence, if students believe they can climb the mountain, even if it is difficult, they are more likely to start down the trail. Like scaffolding, inquiry assessments also motivate students because it involves

them learning knowledge through the active creation of it, not just simply receiving it. Inquiry assessments can be brief prompts that ask students something like; “Should Abraham Lincoln have signed the Emancipation Proclamation?” Questions like this require students to go beyond simple recall of information, they must think about why historical actors committed their actions.

Inquiry assessments can also be large-scale, such as the National History Day projects that thousands of students complete each year. In these projects, students form questions, read sources, and develop a historical argument. Throughout these extended projects, students engage in deeper cognitive processes in order to answer inquiry questions. Inquiring is often more strenuous than receiving a lecture or copying notes, yet inquiry activities that promote the advancement of cognitive processes tend to motivate students more. This is especially true during extended inquiries like the kind National History Day provides.³ Through inquiry, students create knowledge by relating new information to their own perspectives, beliefs, and ideas. It is from forming knowledge through the combination of self, new information, and cognitive processes that motivates students. National History Day, and other large-scale projects, provide opportunities for students to create their own inquiry assessments, thus adding another layer of agency and motivation.

Students being comfortable with topics is key when assessing their historical thinking ability. Students shaping inquiry assessments can provide motivation in an environment where students are uncomfortable or uncertain about the skills they are trying to obtain. For example, at the beginning of a project, students likely will not understand the importance of historical thinking skill like “contextualization”, or what it truly means to detect bias in a secondary

³ National History Day is an organization that sponsors a competition between project-based learning (PBL) style history projects. The organization also promotes a specific type of curriculum for teachers to use.

source. Students need perseverance to understand abstract topics, which teachers can promote by allowing students to shape curriculum and the assessments over their inquiry.

Exemplar Article #1

David Wallace's qualitative analysis of National History Day is an apt research study that demonstrates the power of motivation. Wallace was a professor of history at Cleveland State University in Ohio. In the mid-1980s, National History Day was operated in Cleveland, and Wallace was one of the earliest history professors who helped the program. In 1987, Wallace completed a qualitative research project and wrote the article "The Past as Experience: A Qualitative Assessment of National History Day." The purpose of this research was to describe and evaluate National History Day as an education program, and to describe its implications for teaching history.

Wallace sent a questionnaire to 1,500 students who were state winners in the National History Day program. The questionnaire focused on the perception of the students on how they viewed their cognitive ability because of National History Day. Not only did student participation in National History Day result in increased excitement and engagement with the history curriculum, students believed the program fostered new skills useful for historical research. These skills were evident in the explanations of the students, especially when they described how their theories and evidence involved forces of culture, politics, and economics. The most significant limitation to Wallace's study is the survey was sent only to state winners, who most likely possessed social advantages compared to their less victorious peers. The lack of a more balanced participatory group may skew data since Wallace did not include students who did not move on past the local and state contests.

Wallace's study connects learning historical skills, historical projects, and student motivation to one another. The importance of students obtaining historical skills does not lie within students receiving some score on a test, although there are studies that suggest National History Day, and other extended history projects, do support better standardized test scores (Monaco et al., 2009; Parker et. al, 2013; Sloan & Rockman, 2010). The more important benefit is students creating products that demonstrate their ability to inquire through historical thinking. Finally, completion of the projects motivated students because they were proud of the skills they had learned, and they wanted to demonstrate them outside of their classroom.

Question 5: Why Is It Difficult to Assess Historical Thinking and Inquiry Based Learning?

Although I do not have much space left, I believe it's important to briefly write about this fifth question. It is laborious to measure students' cognitive abilities in history. When considering assessing historical thinking and inquiry, teachers face a lack of instructional time to assess, pressure to satisfy standardized high-stakes tests, and cultural issues with focusing on inquiry over recall/memorization. Monika Waldis's research study elaborates on another problem: the coding mechanisms that assessors need to measure historical thinking are convoluted.

Exemplar Article #1

The use of how narrative changed in quality and structure based on a student's ability to think historically interested Waldis. In order to assess students' historical thinking, Waldis asked German students to produce a narrative based on analysis of primary sources. The study used two topics, "Trade Relations with Japan" and "The Nazi Boycott of Jewish Businesses," which were seldom and often taught respectively in Germany. The teacher gave students exam booklets with primary sources inside of them and gave students as much time as they needed to complete

the analysis questions. The participant sample included 193 high school students from nine classrooms in three different towns. Waldis and the team of researchers developed quality features of narrative assessments, which included concepts such as “value judgments” and “quality of making historical references” (Waldis, 2015, p. 122).

Student answers were structurally heterogeneous; for example, some answers were one sentence long while the longest was thirty-three sentences. The raters of the narrative answers distributed low numbers in the category of *normative cogency* because students did not support their values with evidence and reasoning from the primary sources. This finding is disturbing as it suggests history courses are not educating students to provide evidence to support their own opinions. If teachers are not teaching the concept of using evidence, it could be because the task is more difficult than researchers realized. Waldis used highly detailed coding mechanisms, similar to other research articles but much more complex, and categories when measuring student answers, which resulted in discovering that responses lacked certain qualities. If these complex coding mechanisms are the only way to accurately measure students' acquisition of thinking skills, then teachers will not have the training or background to successfully lead students through historical thinking exercises.

Conclusion

Considering educators did not treat historical thinking seriously until the mid- to late-1990s, there has been an impressive amount of research in how we measure and define historical thinking, on what data researchers can lift from research experiments, and in how inquiry motivates students to succeed. There were even pleasant surprises after I analyzed the studies over historical thinking. Before I examined all of the research, I believed there was an overabundance of focus on high school and undergraduate students in their abilities to

historically think. My position was due to the lesson plans available at historical thinking websites like Beyond the Bubble, SCIM-C, The Library of Congress, and The National Archives, which nearly solely cater to older students. In this literature analysis, I found a variety of research studies that used participatory students in elementary school, middle school, high school, and college undergraduate courses.

Research on historical thinking is incomplete because it is difficult to implement the recommended practices into most of K-16 education. Although researchers have created categories of historical thinking, they have not scaffolded historical thinking skills, nor found ways to simplify them for younger students, or students not at grade level. For instance, different groups of researchers have described how sourcing is an important skill for students to have (McGrew, 2018, Wineburg 2001, Hicks 2008). The assessments I reviewed did not explain sourcing beyond a short definition, along with some complex examples of how teachers can assess sourcing. Teachers will need scaffolded strategies in order for students to work with these skills, such as modeling or simplified versions of sourcing assessments (Shemilt, 2018).

Currently, only a few studies addressed scaffolding historical thinking and how assessment can be tailored to meet student needs. Researchers examining students from multiple age groups gives me hope that researchers will soon conduct studies that examine the relationship between scaffolding practices and the degree to which students can master historical thinking. There are two potential reasons why such studies may not be forthcoming.

First, there simply has not been a lot of time for the inquiry models and beliefs to seep into the education system. For example, the unifying standards document, the C3 Framework, was published in 2013, only seven years ago. For all their goals, the authors of the C3 Framework mainly wanted to create a document that could help social studies departments

vertically align their classrooms based on inquiry. Second, the coding mechanisms, assessment rubrics, and overall process of measuring historical thinking is arduous. Teaching students how to historically think is difficult. Without context, students cannot base their development of questions or their conclusions on social realities of history, economics, geography, or whatever teachers use as curriculum.

Final Questions and Future Research

When considering questions for future research, I wonder what methods of scaffolding historical thinking have researchers tested and what were the results? Based on those results, what additional scaffolds do teachers need? My principal assumption is scaffolding should start with general education and then branch out to other areas. For example, a teacher assessing a specific historical thinking skill, like sourcing, could use alternative resources depending on students' reading levels. One way of doing this is alternating the reading Lexile level of certain primary source documents, which will enable students with a lower reading level to access the key information of the document. These students would then be less distracted by the words that are no longer prevalent in modern language and can better show their analysis abilities in identifying the author and detecting their bias. Finally, there seems to be many confounding factors, like familiarity with a topic, that can skew data of mastering historical thinking skills. Researchers must identify and control these confounding factors in order to get more accurate results.

Appendix A

Sourcing can be classified as a skill that focuses on the analysis of a document or artifact's creator and the circumstances of its creation. There are sub-skills to this type of inquiry which are listed below.

Skill #1: Identified the category of the source.

Skill #2: Identified the date and creator of the source.

Skill #3: Identified if the source is primary or secondary.

Skill #4: Described the audience of the source.

Skill #5: Described the purpose of the source.

Skill #6: Described characteristics, bias, or perspectives of the source's creator.

Skill #7: Described the trustworthiness of the source.

References

- Attewell, Paul and David Lavin (2011). "The Other 75%: College Education Beyond the Elite." In E. Lageman's and H. Lewis's (Eds.) *What is College For? The Public Purpose of Higher Education*. New York: Teachers College Press.
- Bruner, Jerome (1960). *The Process of Education*. Cambridge: Harvard University Press.
- Corliss, Stephanie B., and Marcia C. Linn (2011). Assessing Learning from Inquiry Science Instruction. In D. Robinson and G. Schraw (Eds). *Assessment of Higher Order Thinking Skills*, 219-243.
- Hamblin, Taylor S. [History Forge]. (2019). *Mock Senate Explanation* [Video] YouTube. Retrieved from <https://www.youtube.com/watch?v=7I-ADVLzz4s&t=776s>
- Hicks, David, and Peter E. Doolittle (2008). Fostering Analysis in Historical Inquiry Through Multimedia Embedded Scaffolding. *Theory and Research in Social Education*, 36(3), 206-232.
- Lee, John (unpublished chapter). *Assessing Inquiry*. In the author's possession.
- Lee, P., & Ashby, R. (2000). Progression in Historical Understanding Among Student Ages 7-14. In P.N. Stearns, P. Seixas, and S. Wineburg (Eds.), *Knowing Teaching & Learning History: National and International Perspectives*, 45-94. New York: New York University Press.
- Levesque, Stephane and Penney Clark (2018). Historical Thinking: Definitions and Educational Applications. In S. Metzger's and L. Harris's (Eds.) *The Wiley Handbook of History Teaching and Learning*, 119-148. Hoboken, NJ: John Wiley & Sons.
- McGrew, Sarah, Joel Breakstone, Teresa Ortega, Mark Smith, and Sam Wineburg. (2018). Can Students Evaluate Online Sources? Learning from Assessments of Civic Online Reasoning. *Theory & Research in Social Education*, 46(2), 165-193.
- Monaco, Giuseppe, Bo Lu, & Megan Wood (2009). Impact of the National History Day in Ohio Program on Students' Performances: Pilot Evaluation Project. *Journal of Museum Education*, 34(1), 79-96.
- National Council for the Social Studies (NCSS). *The College, Career, and Civic Life (C3) Framework for Social Studies State Standards: Guidance for Enhancing the Rigor of K-12 Civics, Economics, Geography, and History* (2013). Silver Spring, Maryland: NCSS.

- Parker, Walter C., Jane Lo, Angeline Yeo, Sheila Valencia, Diem Ngyuyen, Robert Abbott, Susan Nolen, John Bransford, Nancy Vye (2013). Beyond Breadth-Speed Test: Toward Deeper Knowing and Engagement in an Advanced Placement Course. *American Educational Research Journal*, 50(6), 1424-1459.
- Reich, G. A. (2009) Testing historical knowledge: Standards, multiple-choice questions and student reasoning. *Theory & Research in Social Education*, 37(3), 325-360.
- Renn, Kristen and Robert Reason. "Characteristics of College Students in The United States". In K. Renn's and R. Reason's *College Students in the United States: Characteristics, Experiences, and Outcomes*, 3-27. San Francisco, CA: Jossey-Bass.
- Shemilt, Denis (2018). Assessment of Learning in History Education: Past, Present, and Possible Futures. In S. Metzger's and L. Harris's (Eds.) *The Wiley Handbook of History Teaching and Learning*, 449-472. Hoboken, NJ: John Wiley & Sons.
- Selwyn, Doug (2014). *Why Inquiry?* In E. Ross's *The Social Studies Curriculum*, 267-288. New York: State University Press.
- Sloan, Kay & Saul Rockman (2010). National History Day Works: Findings from the National History Program Evaluation. Rockman et al.
- Smith, M., Breakstone, J., & Wineburg, S. (2019). History Assessments of Thinking: A Validity Study. *Cognition and Instruction* 37(1), 118-144.
- Stanford History Education Group (n.d.). The First Thanksgiving. Retrieved from <https://sheg.stanford.edu/history-assessments/first-thanksgiving>
- The Glossary of Education Reform (2015). Scaffolding. Retrieved from <https://www.edglossary.org/scaffolding/>
- VanSledright, Bruce (2011). *The Challenge of Rethinking History Education: On Practices, Theories, and Policy*. New York: Routledge.
- VanSledright, Bruce (2014). *Assessing Historical Thinking & Understanding: Innovative Designs for New Standards*. New York: Routledge.
- Virginia Tech. *SCIM-C: Historical Inquiry*. Retrieved from <http://www.historicalinquiry.com/>.
- Waldis, Monika, et al. (2015). Material-Based and Open-Ended Writing Tasks for Assessing Narrative Among Students. In K. Ercikan and P. Seixas (Eds.), *New Directions in*

- Assessing Historical Thinking* (pp. 117-131). New York: Routledge.
- Wallace, David Adams (1987). The Past as Experience: A Qualitative Assessment of National History Day, *The History Teacher*, 20(2), 179-242.
- Wineburg, Sam (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, 83(1), 93-87.
- Wineburg, Sam (1997). Beyond “breadth and depth”: Subject matter knowledge and assessment. *Theory into Practice*, 36(4), 255-261.
- Wineburg, Sam (2001). Picturing the Past. In *Historical Thinking and Other Unnatural Acts: Charting the Future of Teaching the Past* (pp. 113-136). Philadelphia: Temple University Press.
- Wineburg, Sam (2009). Tampering with history: Adapting primary sources for struggling readers. *Social Education*, 73(5), 212-216.

Language and Power

Alessia Barbici-Wagner

Department of Teaching, Learning, and Teacher Education

University of Nebraska-Lincoln

Abstract

The following critical analyses result from different readings related to the idea of the power of language for a graduate level course at the University of Nebraska-Lincoln, which was designed to understand the relationship of language, discourse knowledge and action both in social theory and in educational practice.

During this course we dealt with a variety of issues taken from different disciplines, including but not limited to the following:

- philosophy (Foucault, Butler, Nussbaum, Dewey) –e.g., language as social symbolic power;
- sociology (Bourdieu, de Certeau, Goffman, Wittgenstein, Weber) --e.g., language as symbolic capital;
- semiotics (Barthes, Bakhtin) –e.g., language as social semiotic and dialogic action;
- sociolinguistic analysis (R. Lakoff, Scollon, Cameron, Tannen, Nussbaum, amongst others) –e.g., intersection of language and ideology, gender, work.

The aim of these essays was to understand the problematic nature of language and the way it is used to control knowledge, memory, history, and ultimately, our destinies.

doi: 10.32873/unl.dc.ne009

‘Ketchup’ with Social Norms

That’s not what I meant (Tannen, 2011) and “Is there any Ketchup, Vera?” (Cameron, 1998) treat the same concept of how men and women metaphorically speak a different language, but the authors end up giving two different reasons for it. Even when women and men are part of the same cultural status and share a similar social context, this miscommunication might happen. Most of us grow up with the idea that the key to understanding relationships is through conversation. Now we might doubt that conversation can actually save a relationship, but instead complicate it more.

It is not uncommon to hear people complaining about their relationship with a partner that does not understand them and can lead to a breakup of long-term relationships. Why did this miscommunication between people that live in the same context and culture happen? According to Tannen (2011), it is a problem of misunderstanding tied to the difference in gender and/or the gender hierarchy that our society created. According to Cameron (1998), it is not a misunderstanding but rather a conflict between the two genders. This conflict is explained through the distinction of the male-female gender category, where assumptions — about social roles, positions, rights and obligations — mean a great deal.

Male and female miscommunication has become a myth in our society. There is a general idea that women communicate less directly than men; they are more ambiguous due to a lack of confidence. Cameron’s example of two co-workers, one male and one female, in which the female asks, ‘Where’s your coat?’ and the male answers, ‘Thanks, Mom,’ shows how there are different assumptions in play (Cameron, 1998, p. 440), as well as deductive strategies. The woman in question was hurt. Her friendly suggestion was misunderstood, and the man gained power—he puts her down responding to what he interpreted as a negative remark. He used a

deductive strategy dictated by his utilitarian discourse because, in his view, there's no need to be polite and establish a relationship with a coworker (Scollon, R., Scollon, S., & Rodney, 2012). This example also shows how some misunderstandings are tactical (de Certeau, 1984) by pretending not to understand the real meaning to assert his own position.

The context of the conversation is also a really important element that should not be underestimated. As Fairclough points out, we “should know the social and cultural goings-on which the text is part of” (Machin, 2008, p. 63) and the same applies for relationships when applying Critical Discourse Analysis. According to Machin (2008), these goings-on include “institutionalized habits, procedures, values and the way these are deeply influenced by financial matters” in the same way social and power relations are part of our backgrounds. These constitute our ‘habitus’ (Bourdieu, 1991) and guide our interactions. It might be true that men and women have a different view resulting from their different positions in society, mostly due to our habitus inculcated in young age and which is also part of our culture. And “culture, after all, is the construction of shared meaning” (Lakoff, 2000).

This aspect is more evident in the example about the Ketchup (Cameron, 1998), where during a family dinner the father/husband asks the question “Is there any Ketchup, Vera?” not implying to know if he needed to buy some but with the implicit request or command to Vera to fetch the ketchup for him. In this example, we know that the wife got up from the table to go take the ketchup for her husband because of the social roles and relations that apply in that context. “Language has the means and the medium by which we construct and understand ourselves as individuals, as coherent creatures, and also as members of a culture, a cohesive unit” (Lakoff, 2000). Vera has a sense of obligation towards her husband inculcated as a habitus by the social and cultural context where she lives. The same could not be towards her children as Cameron

(1998) tells us. Had her children asked the same question of Vera, she could use a different strategy answering to her daughter that the ketchup is in the kitchen, assuming that her daughter goes to pick up her own condiment. This is applicable as long as there are the right contextual conditions.

Conflicts between men and women rely on the normative positioning of participants in interactions – “whether in interpreting utterances they make use of conflicting assumptions about the position a particular speaker in a given situation either is, or ought to be, speaking from; and thus hold conflicting beliefs about the right and obligations that are normative in the speaker hearer relationship” (Cameron, 1998) – but this does not eliminate the possibility that women and men can also misunderstand each other without any conflict taking place. “Language is not ‘just words.’ It enables us to establish ourselves, as individuals, and as members of groups; it tells how we are connected to one another, who has power and who doesn’t... Now more than ever language is construed as something worth fighting for, or at least over.” (Lakoff, 2000, p. 41). Language also creates power relations. It is important to underline that this conflict that Cameron (1998) mentions is not seen as positioning women and men on opposite sides, but it is considered a conflict of interest.

These conflicts and misunderstandings result from the myth we created that men have power and women have a subordinate role. We start forming this myth through socialization at birth (think gender colors of baby clothes), and parents support and encourage those roles. For example, parents give toys such as cars, trucks, superheroes, toy guns and more to boys – symbols of strength, masculinity and action that often stimulate motor skills. Girls are often given dolls and dress-up apparel that foster nurturing, social proximity, and role playing. Children then reinforce this socialization through play with the gender-specific toys they are

given. According to Caldera et al.'s study (1998), children will most likely choose to play with "gender appropriate" toys (or same-gender toys) even when cross-gender toys are available because parents give children positive feedback (in the form of praise, involvement, and physical closeness) for gender-normative behavior. Style magazines and, in general, media also help create this myth as in articles where they advise women on how to address their bosses in a less ambiguous way (being more direct is also a prerogative of a strong personality). We should be able to change our habitus in order to change our position in the social context. Moreover, we should be able to change the general assumptions that follow women roles in those specific social contexts for a better communicative relationship.

Immigrants and the Pursuit of Happiness

American people seem not to understand or try to forget that this nation, the United States, is a nation of immigrants. Immigration is an integral part of American history that can be divided into four epochs that brought distinct and diverse national ethnicities and races to the United States: the colonial period, the mid-19th century, the start of the 20th century, and post-1965. After the 9/11 terrorist attack to the nation, and most recently after the election of our current president, immigration became synonymous with 'danger' to our pursuit of happiness and our comfort zone. According to the Immigration Policy Institute (MPI) website (Batalova, J., 2020), about 26 percent of children have immigrant parents, living their lives mostly in in-betweenness – between two different cultures and worlds.

Historically, the United States has welcomed immigrants from all parts of the world, and it was considered a cultural melting pot. Nowadays, the growing concern related to immigrants is in part due to the negative influence of the media, which through their use of metaphors and metonymies do nothing but increase and inculcate negative stereotypes about immigrants. After

9/11, questions such as “What are you?” or new linguistic forms like “being wanded” (Sarroub, L. K., 2002) —referring to the action of the handheld metal detector at the security gates of airports—were minted, while nowadays immigration from Mexico is referred to as a flood. In the media, unaccompanied children of immigrants – who should awaken our empathy— are dehumanized through anti-immigrant discourses and the use of metaphors that describe them as a flood, detainees, criminals and, in some cases, animals.

In pursuing their happiness, Americans see these new individuals as a threat to their way of life. As Ahmed S. points out in her book *The promise of happiness* (2010), “multiculturalism is what makes people unhappy,” forcing people from an outside group to be integrated into an already established group and ruining the existing harmony. According to Ahmed (2010), to make people happier means to make societies more cohesive. The only solution seems to be to “put glue back into communities.” Here “happiness is imagined as social glue.” To achieve this, immigrants should be reoriented to American norms, values, and practices. According to Bourdieu (1991), non-dominant groups will need to adapt to acquire new cultural capital or habits to thrive in the new society, even if they maintain the habitus of their cultural communities.

Nevertheless, we witness that immigrant children that attend American schools tend to find themselves in a space “in-between;” in a state that may not satisfy the expectation of the community where they live. Especially after the 9/11 national terrorist attack, most Arab American communities had to re-present themselves and answer the question “So, what are you?” (Sarroub, L. K., 2002) in order to survive. They live their lives in a constant Foucauldian panopticon, where they are at the center of the observation and racial judgment.

Even TV talk shows don't give them social justice. A recent study of Jay Leno of NBC's *The Tonight Show* (Ana O. S., 2009) showed and analyzed the anti-immigrant jokes' effects on a national television audience in 2006. These shows together with news and other media do nothing but feed the anti-immigrant feelings by promoting intolerance. Through those jokes, Jay Leno was able to give relief to his audience which was assisting at the political Great May Day Marches of 2006⁴, since jokes tend to address topics that carry emotional weight for the audience. According to Ana's (2009) article, when someone laughs at those jokes he/she feels "asset superiority over someone who we deem to be our inferior." In his jokes immigrants become beasts, prostitutes and a national danger when, for example, he joked that "Mexicans have so many children that they will soon overrun the United States," he is able to dehumanize the immigrants and to underestimate their effort to find happiness. When people laugh at his jokes, they align themselves with the comedian while distancing themselves with the subject of the joke (the butt of the joke.)

A better utilitarian ethical solution to the "problem" of illegal immigration that might help all of us pursue our happiness without racial or ethnic distinction would be to reform the naturalization process for U.S. citizenship. As part of the naturalization process, applicants for U.S. citizenship must pass a two-part naturalization test. The first component is an English test that assesses the applicant's ability to read, write, and speak in the language. The second, a civic test, evaluates the applicant's knowledge of U.S. history and government. Unfortunately, this process is discriminatory because it is only offered in the English language despite the fact that

⁴ In 2006-2007, millions of people participated in protests over a proposed change to U.S. immigration policy. Great May Marches of 2006 estimated 5 million people marched in more than 100 cities across the country (Ana O. S., 2009)

there is no official language of the United States and that there are tens of millions of people living in the U.S. that speak a language other than English.

In *Creating Capabilities*, the author Nussbaum (2013) argues that we need to refocus our idea about development on the scale of individuals. Development of a society and its happiness is not about how rich the nation is, but rather it is about whether people can live in a way worthy of human dignity. We cannot expect immigrants to abandon their culture and their language, but rather we should think that those will enrich ours.

Culture Turns Deadly

The general nervousness and doubts that an airplane passenger typically has before boarding the aircraft is not alleviated or made easier by reading the seventh chapter of *Outliers* (2013). In this era, where distances among countries are made shorter by the use of planes as a means of faster transportation, it seems that ethnic differences can instead cause a distance issue for dialogue among the plane crew and eventually lead to the crash of the plane. The same question that Gladwell (2013) raises: "Why is the fact that each of us comes from a culture with its own distinctive strengths and weaknesses, tendencies and predispositions, so difficult to acknowledge?" (p.221) is a question that people in general should ask themselves, but especially colleagues in a workplace. Bourdieu (1991) says that cultures are part of the baggage of habitus that grows inside of us and makes us who we are. It would seem that according to this idea we are able to change our cultural habitus and avoid such incidents.

In this chapter, Gladwell (2013) analyzes the reason why plane crashes happen and, astonishingly, he realizes that this is not always an engine malfunction, weather or personal failure, but a more serious and catastrophic failure to understand and acknowledge cultural differences. According to R. & S. Scollon (2012), there are four major factors in intercultural

communication: 1) Ideology formed by the history and the worldview such as beliefs, values and religion; 2) Socialization; 3) Form of discourse (such as the function of language, the non-verbal communication) and; 4) Face system, meaning a social organization not only inside the family and the community, but also in the society where you live. All these factors are important and should not be underestimated regardless of the situation. In one example given in Scollon's (2012) article, these cultural differences between pilot and copilots resulted in an ambiguous communication that led to the plane actually crashing.

Korean Air flight 801 was taking a route from Seoul to Guam and was piloted by an experienced captain who was also familiar with the route. Yet the plane never reached its destination, eventually crashing into the side of a mountain. By the late 1990s, the airline had a terrible reputation for crashing – it had 17 times more crashes per million departures than any American airline. Transcripts of the crashes covered in this chapter show how ambiguous or non-direct communication was a major factor in the crash. “When we ignore culture, planes crash,” the root of such attitudes is cultural.

In this particular case, the respectful speech dictated by hierarchal Korean and Chinese cultures might be appropriate for most situations; however, in the cockpit of an airplane running low on fuel and looking for an alternative route around out of bad weather necessitated clear and direct speech or “transmitted oriented” – a communication that considers the responsibility of the speaker to communicate ideas clearly and unambiguously. A utilitarian communication, as in the sense of Scollon, R. & S.(2012), is a communication that has a purpose and, in this case, would have been useful to eliminate ambiguity.

Cultures seem to be a baggage of habitus; we grow with the idea of hierarchy, respect, individualism and more. Changing that habitus for Bourdieu (1991) seems impossible. At the

end of the chapter, David Greenberg, an outsider from Delta Air Lines, was able to change that cultural habitus, at least in that specific context or “market place.” “Language was the filter” – he did not assume that legacies are an indelible part of the whole like Bourdieu (1991). Instead, he taught those people the English language, thereby giving them an “alternate identity” where they could forget their culture and act accordingly. They transformed their relationships to their new work environment. In that context, they received “education” (as a formal teaching and learning) and were “enculturated” (socialized, or informal teaching and learning, as a way to learn from colleagues or observation), while at the same time “acculturated” (when two different cultures come into contact and the stronger or more powerful influences the other) (Scollon, R. & S., 2012.)

It might appear that the Korean crew was able to use strategies, in a de Courteau way, that allowed them to push away their hierarchy culture only during their work time. Or, in a Bourdieu way, they were able to transform their “self” into another self on that specific occasion. Nevertheless, the issue was resolved and with the introduction of the English language and the American individualistic “transmitter orientation” communication, the Korean “receiver oriented” – where it is up to the listener to make sense of what is being said— was able to be modified to save airplanes and above all lives.

A Positive Perspective for the Children of Immigrants

Immigrant and refugee children in the U.S. have to face many difficulties due not only to the adaptation of their family to their new situation, but also in seeking acceptance by the community in the “white public spaces” where they moved and live or where they were born. They are trying to integrate into the American social and educational system, trying to overcome the ethnic and racial barriers of a white society that are perpetuated by media bias and those who

are in power. These children have to live with their contradictions, simultaneities, and conflicts in order to be accepted and help their family to be accepted and integrated. These children, in their struggle for survival, will end up learning early on that they must sacrifice their own desires for the good of the family.

Those who are the children of immigrants in the U.S. seem to live two lives in limbo between their origins and monolingual families and immersed in their new American culture. Those children who moved here as immigrants or, worse, refugees are trying to adapt themselves to a completely new situation and simultaneously learn a new linguistic capital as well as creating a new habitus that will make them integrate in the new system to become legitimate participants in society. In the “New Immigrant Youth Interpreting in White Public Space” (Reynolds & Orellana, 2009), bilingual children of monolingual immigrants act as interpreters and translators of the language for their parents who have limited capabilities to understand and speak the language. In doing so, these children live with adult responsibilities.

Bakhtin’s (Holquist, M., 2002) concept of dialogism as “the notion that words carry histories and ideologies that frame subsequent interactions as they unfold ontologically” (Reynolds & Orellana, 2009) is important in this context to understand how these young interpreters negotiate the different and often unknown linguistic registers to convey their messages. Bakhtin’s (Holquist, M., 2002) dialogism highlights a dialectic relationship between human beings and social contexts. When they decide the words they are translating or *para*-phrasing for their family, they also become authors of “discourse” that is dialogic since they are negotiating based on their needs and the resources made available to them from their families and society.

At the same time, while providing service and surveillance “within overdetermined interpreter-mediated practice” they live in a state of anxiety due to the task they are asked to do. According to Foucault (1975), they are now the center of a spectacle that sees those little interpreters observed, analyzed and judged by their own monolingual parents and those who are interacting with them in the white public space. Both children who came in the U.S. as immigrants/refugees and those who are from an immigrant family but born in the U.S. will soon learn to use the strategies and tactics, per de Certeau (1984), as a way of operating in the world in order to obtain resources or to avoid inconvenient situations more often related to their unequal positions.

People normally live with the struggle against external definitions or judgment of their thoughts and actions, which produce an effect on them. Those children have to struggle to be part of their new community to help their families survive, but mostly they struggle to find a meaning that will suit both sides of their auditorium (e.g. customers/salesclerk, or students/teacher). They find themselves standing in the middle of a conversation between two languages mediating conversations for their own or their families’ social survival. By engaging mostly with adults in positions of power, these children often conduct dialogues that will shape their perception of reality and their habitus while developing register-specific competencies by forging new roles and identities.

According to Bakhtin (Holquist, M., 2002), we are always in dialogue with others and everything else in the world. Each of us is uniquely situated in a particular place and time in the world. One can see one’s exterior only through others’ perspectives from which one can produce something new or enriching. And this is the positive perspective of those children. In the end,

these children will have more to offer our society than the children that never had to develop those specific strategies or tactics.

Dialogism in the New Technological Era

Living in the new technological era, where dialogue and in-person interactions between people are gradually diminishing in favor of electronic communication, one must ask if the ideas of the main French thinkers such as de Certeau, Bakhtin or Bourdieu would have changed. If the concepts of habitus, dialogism or, in general, the idea of the ordinary language would have been revisited. The relationship between technology, dialogue and the self must be explored in order to analyze how they differ when compared to the era without digital technologies. Do mass media, blogs, and platforms like Facebook, Twitter and Instagram diminish or enhance opportunities for dialogue?

Dialogism, according to Bakhtin (Holquist, M., 2002) is “unthinkable outside its relation to language” and this relation is dialogic or based on the dialogue between people. ‘Speaking’ and ‘exchange’ are essential aspects of dialogue. Bakhtin’s (Holquist, M., 2002) concept of dialogue appears to see speakers in a conversation face-to-face, which seems impossible for this particular era where most of the conversation are actuated by electronic communication.

Electronic communication seems to prioritize information over genuine conversation among humans. We see posts on Facebook that report the latest news or trends where the extent of most interactions is a ‘like’ or an emoji to express dissent or approval. Communication by text messaging seems to have suffered the same reductive effect. The habitus (Bourdieu, 1991) of the people that are using the machine for everyday communication is changing accordingly. They now transform their habitus to adapt new tactics and strategies (de Certeau, 1984). Politicians often use Twitter as a strategy (art of the strong), not only as social recognition, but also as a way

of operating in the world and to gain more power. In particular, Twitter is an effective, low-cost tool of power through which politicians can share messages among their followers, self-promoting and criticizing their opponents/critics. Tactics (art of the weak) are employed by the many who use the internet to steal other people's resources (scammers, identity thieves).

Many perceive the threats of the new interaction through technologies at the expense of one's identity revolving around anonymity. Many psychologists suggest that technology leads to distant human relationships and eventually leads to solitude. Others think that it also can lead to a culture of individualism and narcissism. The Internet becomes the new "panopticon" where participants become the center of the attention and judgment by the onlookers (Foucault, 1975). Consequently, those who post feel the pressure the society places on the content they create. However, all of those views do not seem to support or accommodate genuine dialogue among people.

According to this pessimistic analysis, humans are tempted to succumb to the power of the technology at the expense of human relationships. At the same time, it could be that digital technologies actually have the potential to create a new form of dialogue. Referred to as "telelogic communication," this new type of dialogue occurs between people that communicate electronically.

For example, schools are introducing new digital technologies that expand the dialogue among students. Technologies and pedagogy are starting to work together to improve the educational experience. Google translate is improving every day, helping to translate or communicate with foreign languages. Foreign barriers seem to be getting smaller. Teleconferences can help reach different parts of the globe to interact in a dialogue.

In conclusion, technologies do not eliminate the relationship and the dialogue among humans, but they do change the way it is conceived. Bakhtin (Holquist, M., 2002) talks about a dialogue at the boundaries between inner and outer human experience. You need not be physically present to have a dialogue with someone – technology can assist in this endeavor at either the inner or outer experience.

References

- Ahmed, S. (2010). *The promise of happiness*. Durham, NC: Duke University Press.
- Ana, O. S. (2009). Did you call in Mexican? The racial politics of Jay Leno immigrant jokes. *Language in Society*, 38(1), 23–45. doi: 10.1017/s0047404508090027
- Bourdieu, P., & Thompson, J. B. (1991). *Language and symbolic power*. Cambridge, Mass: Harvard University Press.
- Caldera, Yvonne, Aletha Huston, and Marion O'Brien. (1998). Social interactions and play patterns of parents and toddlers with feminine, masculine, and neutral toys. *Child Development*, 60(1):70–76.
- Cameron, D. (1998). "Is There Any Ketchup, Vera?: Gender, Power and Pragmatics." *Discourse & Society*, vol. 9, no. 4, pp. 437–455., doi:10.1177/0957926598009004002.
- de Certeau, M (1984), *"The Practice of Everyday Life"*. Berkeley: University of California Press
- Foucault, M. (1975). *Discipline and punish: The birth of the prison*. New York, NY: Vintage Books.
- Gladwell, M. (2013). "The Ethnic theory of plane crashes." From *Outliers: the story of success*. New York: Back Bay Books, Little, Brown and Company.
- Holquist, M. (2002). *Dialogism: Bakhtin and His World*. London. Routledge.
- Jeanne Batalova, J. (2020, March 31). Frequently Requested Statistics on Immigrants and Immigration in the United States. Retrieved August 06, 2020, from <https://www.migrationpolicy.org/article/frequently-requested-statistics-immigrants-and-immigration-united-states>
- Lakoff, R. (2000). "Language: The Power We Love to Hate," from *The Language War*, Chptr. 1, pg. 17-41; Berkeley: University of California Press
- Lickerman, A. (2010). The effect of technology on relationships. Retrieved from <https://www.psychologytoday.com/us/blog/happiness-in-world/201006/the-effect-technology-relationships>

- Machin, D. (2008). "News Discourse I: Understanding the Social Goings-On Behind News Texts," from *Language and Power: An Introduction to Institutional Discourse*, Chptr. 4, pg. 62-89; London: Continuum
- Nussbaum, M. C. (2013). *Creating capabilities: the human development approach*. Cambridge, MA: The Belknap Press of Harvard University Press.
- Reynolds, J. F., & Orellana, M. F. (2009). New Immigrant Youth Interpreting in White Public Space. *American Anthropologist*, 111(2), 211-223. doi:10.1111/j.1548-1433.2009.01114.x
- Sarroub, L. K. (2002). From Neologisms to Social Practice: An Analysis of the Wandering of America. *Anthropology Education Quarterly*, 33(3), 297–307. doi: 10.1525/aeq.2002.33.3.297
- Scollon, R., Scollon, S. & Jones, R. (2012). *Intercultural Communication: A discourse approach*, 3rd edition. London: Wiley-Blackwell.
- Tannen, D. (2011) *That's Not What I Meant! How Conversational Style Makes or Breaks Relationships*. Harper.

Seeing Formative Assessments from a Broad Perspective

Consuelo M. Gallardo

Department of Teaching, Learning, and Teacher Education

University of Nebraska-Lincoln

Abstract

Despite the fact that many experts in the assessment field have advocated for the use of formative assessments, little attention has been paid to their thorough elaboration and application in Ecuadorian English as Foreign Language (EFL) classrooms. Some teachers have overlooked the validity of formative assessment as tools to inform instruction and language learning growth, so its application has generated a big debate. Therefore, this paper presents a literature review of perceptions and experiences of diverse scholars and practitioners who have analyzed the validity of formative assessments and conducted studies in EFL classrooms. It starts by providing information related to the characteristics and purposes of formative assessments. It also describes their application and impact on teaching and learning. It highlights how formative assessments can invite students to take learning ownership and provide immediate feedback to improve language development. Subsequently, it shows the advantages and challenges of formative assessments, which include peer reviews, peer assessments, self-assessments, and portfolios. Additionally, it points out the effects of formative assessments to check students' language proficiency growth and teacher awareness. Finally, it discusses a summary of encouraging and challenging literature review findings, future research questions, and a call for action. This literature review invites EFL teachers to see formative assessments from a broad perspective increase their awareness and reflect on their further applicability.

Keywords: teaching, formative assessment, feedback, English language skills

doi: 10.32873/unl.dc.ne010

Over the years, teachers have been required to conduct standardized tests for English language learners to assess their English knowledge, as well as their receptive and productive skills. This summative assessment practice has been conducted in English-speaking countries and in countries where English is considered as a second or foreign language. Although summative assessments have been considered as reliable tools to gather data that informs instruction and helps learners improve their English proficiency, it is not necessarily the case in some Ecuadorian EFL classrooms.

To illustrate, as an English teacher I have heard high school students complaining about summative assessments and stating that they have not been very useful to advance in their English level. Some students have even argued that these assessments do not have consequential validity since they do not see them as learning experiences. Besides, they usually emphasized that summative assessments do not allowed them to receive effective feedback in order to improve their English proficiency level. As Pat-El, Tillema, and Segers (2013) mentioned, it is difficult for students to obtain real evidence of their learning progress and reach high standards if they only receive a quick and general review based on their final tests' results. It leads teachers to move on without actually reinforcing students' language skills and make some learning improvements.

In the Ecuadorian case, while working as a teacher, I have seen that not only students expressed dissatisfaction with their lack of English proficiency, but also educational authorities had concerns about students' low English level. They have argued that although in 2011, the Ecuadorian national guidelines from the Ministry of Education established that students have to take five English class periods per week throughout six academic years, students do not have an intermediate English level by the time they graduate from high school. Consequently, they

complain that students have not acquired the necessary knowledge and skills to reach the mandatory *Intermediate English level (B1)* based on the *Common European framework of reference for languages* which have been adapted from the Council of Europe (2011). As a result, this reality has forced higher education authorities to continue hiring English teachers so that college students can take more courses for three or four years until they reach the intermediate English level.

Besides, Serrano, Vizcaíno, Cazco, and Kuhlman (2015) indicated that many Ecuadorian English teachers do not have “the methodologies to teach English effectively in the schools” (p. 109). As such, if research shows that English instruction has been ineffective and not good enough to help EFL students, it is crucial to start considering not only useful teaching practices but also assessment tools that allow teachers and students to obtain better English language achievements.

As Black and William (1998), Bloom (1969), Cauley and McMillan (2010), as well as Hattie and Timperley (2007) highlight, formative assessments are tools that assist teachers to inform their instruction and provide the necessary support to help students improve their English language knowledge and skills. Besides, they argue that formative assessments report valid information about students’ language progress and allow teachers to provide constant and immediate feedback since these assessments reveal specific learning strengths and weaknesses. Hence, this supportive assessment practice allows students to receive scaffolding instruction to advance them in their learning and be better prepared to get high achievement standards.

Despite the fact that many experts in the assessment field have advocated for the use of formative assessments in EFL classrooms, little attention has been paid to their thorough elaboration and application in many countries, including Ecuador. To exemplify, while working

at the public and private sector, I had observed that some colleagues do not consider formative assessments as practical tools since not only its elaboration, but its application demands too much work, especially when having forty or fifty students per class. Thus, formative assessments have been seen as tools that have unclear grading purposes, demand a lot of planning and grading, and reduce class time to complete other grade-level contents.

Consequently, there is a need to explore more theories and research studies like the one from Lee and Coniam (2013) about the effective use of formative assessments in EFL classrooms to advocate for their effective implementation in public high schools. Once teachers realize that formative assessments can help them gather real evidence about the students' language progress and inform their instructional practices, they would start relying on them. What is more, they would be more likely to start applying them in their classrooms when assessing students.

To accomplish the overarching goal of inviting EFL teachers to see formative assessments from a broad perspective, this research-based paper presents perceptions and experiences of diverse scholars and practitioners. It starts describing information about formative assessments. It also highlights the impact of formative assessment in EFL classrooms and its implications and presents information about the relevance of teacher training before implementing them. Finally, it closes with a call for action.

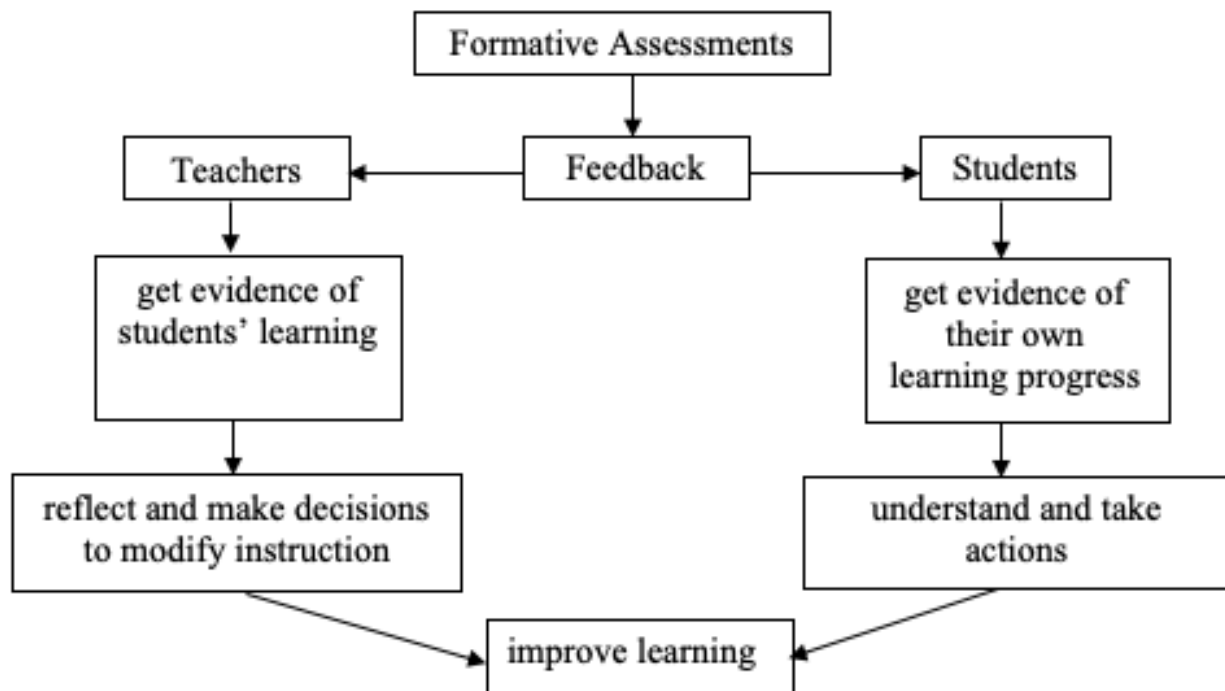
Literature Review

Dynamic Assessment Characteristics

Assessments are considered as important teaching and learning tools since they influence both teachers' and students' lives. Thus, according to Richards (2015), their application has been

essential not only to inform teacher instruction but to monitor student progress while identifying strengths and weaknesses to providing effective feedback.

Varying Purposes of Formative Assessment. When describing formative assessments or also known as assessments for learning, Brown and Abeywickrama (2018) state that “they have as their primary focus the ongoing development of the learner’s language” (p. 8). It allows teachers to constantly assess student progress that occurs during the process of an instructional cycle. Similarly, these assessments gather relevant data of student learning to provide support to modify instructional practices. Gottlieb (2016) also affirms that formative assessments allow teachers to receive permanent and descriptive feedback on students’ knowledge and skills for making improvements to help them advance in their learning. Likewise, Gordon and Rajagopalan (2016), Heritage (2010), Moss and Brookhart (2009), and Popham (2008) have mentioned that the main purpose of formative assessments is to offer relevant information (feedback) for teachers and students. On one hand, teachers use formative assessments to get real evidence of student learning, go under a reflective process, and make decisions to modify their instruction to facilitate student learning. On the other hand, students receive specific evidence of their own learning progress, understand and identify their strengths and weaknesses, and take actions to keep improving. Despite the fact that many authors have provided varying functions of formative assessment as tools that offer information about the teaching practice and learning growth, Figure 1.1 attempts to summarize its most common purpose.

Figure 1.1*Formative Assessment Purpose*

As assessment practices are fundamental for identifying student performance, Gottlieb (2016) emphasizes that it is vital to have specific purposes when assessing English language students in order to avoid misinterpretation of their results. To exemplify, if an assessment purpose is to determine whether or not students need extra language support, teachers can apply language and literacy surveys. When the purpose is to enhance English language teaching and learning, teachers can implement instructional assessments, peer assessments, and self-assessments. However, if its purpose is to monitor students' language progress over time, teachers can use portfolios and journals. Thus, the following information highpoints research exemplars to show how formative assessments can be useful as long as there is a clear purpose for their implementation.

Essential Aspects about Formative Assessments. When conducting formative assessments, it is relevant that teachers have a clear idea of what they want to assess, and how, when, and why they want to do it. This leads to formulate clear purposes and establish well-defined descriptors so that students can show their learning while teachers take time to unveil their strengths and weaknesses related to content and language skills.

To exemplify, Chappuis and Stiggins (2002) mention that formative assessments are seen as effective tools by students when they have a clear idea of what it is the expected overarching learning goal that they need to accomplish, what their current and future learning achievement looks like, and what necessary actions they should take in order to advance in their learning. To provide more information about it, three essential aspects of formative assessment practices are depicted in Figure 1.2.

Figure 1.2*Essential Aspects of Formative Assessment Practices*

Formative assessment practices...	Students need to...
offer students a clear final learning goal they need to accomplish and explains why they need to do that	know the final learning goal
provide students specific information of their current knowledge and skills in comparison with their final learning goal	understand the current learning
motivate students to take actions to improve their knowledge and skills in order to reach the final learning goal.	meet the final learning goal

This information clearly explains that formative assessments provide concrete and useful data for both, students and teachers. While students use the provided data to monitor their own learning growth, teachers can use it to differentiate their instruction in order to meet students' needs. As Gottlieb (2016) highlights, students and teachers play an important role when assessing learning. Therefore, there is no reason for teachers to feel overwhelmed and think that they have all the responsibility to conduct multiple formative assessment practices by themselves.

Incorporating Formative Assessment in EFL Classrooms

In some EFL classrooms, students have been part of traditional assessment practices for many years since teachers have relied on standardized test to check students' English knowledge. However, as Colley (2008) highlights, traditional assessment practices that only takes into account standardized tests to evaluate students do not really show what students know. Besides, they do not invite students to take ownership of their own learning, and it leads students to have constant struggles to see clear English learning improvements. Therefore, when students do not have the opportunity to take part in their language assessment process, so they just rely on teachers to receive feedback and guidance to improve their language knowledge and skills.

Nevertheless, not only Bloom (1984) but also Leahy, Lyon, Thompson and Wiliam (2005) mention that teachers can use formative assessments as tools to invite students to take more responsibility for their own development throughout the learning process, and start relying more on their classmates' support.

Promoting Ownership for Learning. There have been several researchers who have conducted studies in EFL classrooms to see if formative assessments are reliable methods that can make a positive impact and give students the opportunity to take control of their own learning. For instance, Jing Jing (2017) conducted a case study in Hong Kong and included one teacher and eight students in order to examine whether or not the use of formative assessment practices in an EFL classroom can facilitate student self-regulation. Thus, the use of lesson observations, video recording, teacher and student interviews, and surveys allowed the researcher to find out that formative assessment practices really support student self-regulation. Similarly, this research findings showed that these assessments help students clarify their understanding, receive immediate and specific feedback, and facilitate their self-assessment process.

Although the results from Jing Jing's (2017) study indicated that students' perceptions of formative assessment practices are positive, they also reveal that students appreciate more when getting feedback from their teachers rather than from their classmates. It clearly shows that EFL teachers need to keep on using these assessment practices in English language classrooms and help students become more familiar with the feedback process. It can lead them to appreciate and rely more on their peers' suggestions. If students are frequently engaged in assessing each other, they can start seeing the validity of both, their teacher's assistance and their peers' support.

As Jing Jing (2017) mentions, students' lack of trustworthiness on formative assessment practices like peer assessment is due to their fixed perception of teachers' and students' roles in the teaching and assessment process. This shows that there is still a lot to be done in order to encourage students to change their negative perceptions towards peer assessment and feedback and move from a teacher-center to a student-center scaffolding. If more EFL teachers use formative assessments like peer assessments to encourage students to value their classmates' contributions, they might start taking more ownership for their learning in order to improve their English knowledge, as well as their language skills.

Improving Writing Language Skills. Assessing language productive skills like writing may be a challenging job, especially when working in EFL classrooms that have many students. Therefore, it is vital to present information about how formative assessments have been increasingly incorporating into several EFL classrooms around the world as a way to promote student learning and participation, as well as to decreased too much teacher intervention.

For instance, one study conducted by Naghdipour (2017) has analyzed the application of formative assessment in the Iranian EFL classroom and their impact on students' writing performance and their attitudes toward writing. The researcher analyzed thirty-four, first-year

undergraduates and used pre- and post-study writing tasks, pre- and post-study questionnaires, and semi-structured interviews to collect information about their performance. The results of this study show that the application of formative assessment helped students improve their writing skills and develop positive attitudes toward writing.

In addition, the findings described that students value having formative assessment strategies, such as peer review, peer- and self- assessments, as well as portfolios since these tools let them identify specific strengths and weaknesses to enhance their writing tasks. As a result, this study reinforces the idea of implementing formative assessments to foster EFL student accountability for their learning, reduce teachers' workload by having students assess and support each other. What is more, it these kinds of assessments allow teachers to closely monitor students' improvement and keep a record of their writing progress.

Giving and Receiving Feedback. Despite the fact that formative assessments have been considered as important tools to identify strengths and weaknesses in order to provide immediate feedback on student learning, there may be some EFL teachers and students who do not value them. Thus, it is relevant to examine Burner's (2016) research work which examined the validity of formative assessments from the teachers' and learners' perspectives. This four-year longitudinal study was conducted in Norwegian EFL writing classes in order to gather information about teachers' and students' perceptions of formative assessments. After using teacher interviews, focus-group student interviews, and surveys, Bruner (2016) found out several contradictions between teachers' and students' perceptions of the formative assessment of writing. These contradictions were related to feedback, grades, text revision, self-assessment, and student involvement.

For instance, Bruner's (2016) reported that when students were exposed to formative assessment practices, they did not really see their value since teachers just provided negative comments. Nevertheless, teachers affirmed that they have provided as much effective feedback as possible in English and Norwegian so that students can improve their writing skills. This clearly shows that even though formative assessments allowed teachers to provide feedback, it was not appropriate for students since teachers just concentrated on the negative aspects.

To avoid having these contradictory perspectives about receiving and providing feedback, it is essential to evaluate whether or not the provided comments are useful. As Black and William (1998) underline, the use of feedback needs to be considered as a tool to scaffold learning since it enables students to receive specific information about how to improve particular skills instead of just pointing out something that is incorrect.

Apart from these contradictions about the feedback generated by the application of formative assessments, Burner (2016) also reports some challenges. One of them showed that applying formative assessments to check students' writing progress was time-consuming for teachers since it requires to provide individual and detailed written feedback for each student. This study findings also revealed that student were frustrated after receiving feedback from their teachers since they had lack of school time to have one-on-one meetings to follow up and discuss on their teacher comments. Together, the contradictions and challenges based on teachers' experiences and students' perceptions about formative assessments point out that before advocating for their implementation, it is extremely necessary to make sure that teachers and students have a clear understanding of how to give and receive formative feedback to make some sounded contribution and progress.

Monitoring Oral Performance. Assessing students' English oral production when using formative assessments has been described as an arduous task for EFL teachers (Ruiz-Primo, 2011). As such, it is relevant to analyze research studies to see how other EFL teachers have used them, as well as to examine students' reactions towards these assessment practices. To illustrate, Hirasawa (2013) analyzed five English teachers and twenty EFL Japanese students in a foreign language institute in Tokyo in order to see the effect of self- assessment on foreign language students' oral skills. The study lasted three months and let the researcher conducted pre- and post-tests, and videotaping interviews to gather important data about formative assessment applications and students' reactions.

For this study, Hirasawa (2013) used a self-assessment checklist with subcategories related to the organization of the presentation, content, linguistic factors, and interaction with the audience. The major findings indicated that using self-assessments have a positive influence on students' oral performance ability since they have the opportunity to reflect on their own learning growth. Although the results were positive, the researcher suggests analyzing and interpreting them carefully, since the study had a small number of participants.

This study shows that despite the fact that there were positive findings of formative assessments as tools to help EFL students develop independent strategies to recognize their strengths and weakness to improve their oral production, there is the need to conduct more research studies taking into account a larger population to validate the results.

Exploring General Language Proficiency Growth. Over the last years, it has been difficult to reach an agreement about the implementation of formative and summative assessments in EFL settings. Thus, exploring and presenting information about research studies on the impact of assessment methods on foreign language proficiency growth is essential. For

instance, a longitudinal study conducted by Ross (2005) analyzed and compared language proficiency growth when using formative and summative assessments. The study lasted eight years and included 2215 Japanese EFL undergraduates that belonged to eight cohorts of foreign language learners. This study used a mixed-mode approach, parallel growth models, a group added growth model, and direct multivariate tests. The research results indicated that formative assessment practices, such as self- assessments, peer-assessments, portfolios, and group projects produce fundamental skill-specific effects on language proficiency growth. For instance, the findings show that these assessments helped improve students' listening comprehension, but there was not much academic reading growth. Hereafter, Ross (2005) suggests conducting future meticulous investigation to analyze the influences of this discrepancy of growth across English language skills.

This research study offers information about positive students' language proficiency growth, but it also shows possible limitations of certain formative assessment practices to help EFL students improve their four language skills: listening, speaking, reading, and writing. As such, it reinforces the need for evaluating the impacts of formative assessments on specific English language skills to analyze to what extent certain strategies may or may not help students improve their overall language proficiency.

Unveiling Teacher Understanding of Formative Assessments

Some researchers argue that it has been difficult to identify for certain whether or not teachers use formative assessment in their classrooms as tools to inform their teaching and support students to enhance their English knowledge and language skills. To analyze this issue in depth, researchers like Foster and Poppers (2009) as well as Kuzel and Shumba (2011) have examined teachers' formative assessment practices. These studies explained that most teachers

did not know about formative assessments, and only a few had a general idea about them but were not able to properly design them. As such, students only got a score and did not receive any help to improve their learning achievement.

Observing Teacher Assessment Practices. It has been challenging to determine whether the use of formative assessments really improve learning or if they are simply used as grading tools. Therefore, Widiastuti and Saukah (2017) have conducted qualitative research to analyze teachers' understanding of formative assessments and how they influenced the follow-up actions in response to students' learning problems. The researchers included three junior high school teachers and three junior high school students in their study. The data for this study was collected through semi-structured interviews. The research findings revealed that English teachers had a low understanding of the nature of formative assessment and could not provide effective feedback to help students improve their learning. The results also showed that there was no implementation of proper follow-up actions to enhance the teaching and learning process. All in all, this research work explains the need of teachers to get more information about the role of formative assessments, the use of feedback, as well as the application of follow-up activities to scaffold students and benefit their learning.

Analyzing Teacher Assessment Preparation. In 2017, another study was conducted by Saito and Inoi in order to analyze the use of formative assessment by EFL teachers. The researchers included 727 participants who were Japanese junior and senior high school EFL teachers. This research work used William's (2010) model of formative assessment strategies to examine differences in the use of formative assessments in EFL teaching contexts. This study gathered information through a large-scale survey, interviews, observations, as well as follow-up questionnaires and emails. The results showed varying degrees of formative assessment use and

classified teachers into three categories: high, middle, and low users taking into account four variables: intentions, methods, purposes, and feedback. Although Japanese teachers were classified in different users' categories, the study found that most of them were classified as middle and high users which shows teachers' readiness to apply formative assessments.

Subsequently, this study points out the relevance of providing a training program for teachers so that they can effectively apply formative assessments in EFL classrooms. It also emphasizes that once teachers start implementing formative assessments in their classrooms, it is important to have specific instruments to evaluate them, such as the theoretical framework proposed by William (2010).

All the aforementioned studies emphasize that there are multiple factors strongly influencing teachers' and students' implementation of formative assessments. Consequently, they need to be examined in depth in order to accomplish their overarching function of "evaluating students in the process of "forming" their competences and skills with the goal of helping them to continue that growth process" (Brown and Abeywickrama, 2018, p. 8).

Conclusion

Encouraging and Challenging Findings

The body of literature described above emphasizes the relevance of having a clear understanding of the purpose of formative assessment practices and describes several benefits and challenges of their implementations in EFL classrooms. To start, Gordon and Rajagopalan (2016), Gottlieb (2016), Heritage (2010), Moss and Brookhart (2009), and Popham (2008) have underlined that before implementing these assessments, teachers and students need to know why it is relevant to use them and how to use them. Otherwise, it is going to be difficult for teachers

to provide effective feedback based on the assessment results, and for students to understand it in order to advance in their learning and achieve specific goals.

On one hand, studies from Leahy et al. (2005) and Jing Jing (2017) report that students actually value formative assessments since they help self-assess their learning as well as receive detailed feedback from their teachers. Nonetheless, the common concern of their research shows that students do not really appreciate receiving feedback from their peers.

On the other hand, Burner's (2016) study points out two specific findings. The first one is that students do not value receiving feedback from their teachers since it only underlines negative aspects. The second one shows that providing written feedback is challenging since it is time-consuming for teachers, and students also report not having plenty of class time to discuss and analyze that kind of extensive feedback with their teachers.

Additionally, when analyzing the influence of formative assessments to improve language skills, Naghdipour's (2017) work points out that peer reviews, peer- and self-assessments, and portfolios help students improve and develop positive attitudes towards writing. Hirasawa (2013) also explains that when formative assessments are properly applied, they allow students to identify their strengths and weaknesses related to their oral performance. However, Ross' (2005) study shows that these assessments need to be carefully applied and have an impact analysis on specific language skill improvements. For example, her research results report that formative assessments help students improve their listening comprehension, but they do not support their academic reading growth.

Furthermore, the research findings and analysis of Foster and Poppers (2009), Kuzel and Shumba (2011), as well as, Widiastuti and Saukah (2017) explain that when teachers do not have a clear understanding of formative assessment practices, it is difficult for them to design and

provide useful follow-up activities to enhance EFL students' learning. Therefore, as Saito's and Inoi's (2017) highlights it is relevant to know about formative assessments' methods and purposes, as well as, how to provide feedback to the learners before implementing them in class. Their research study reveals that the majority of EFL teachers, who knew about formative assessments, were able to properly apply them in their classrooms.

Future Research

The inconsistency of the research results previously summarized emphasizes the need to conduct future research to examining not only teachers' knowledge of formative assessments but also their implementation and students' reactions toward them. Thus, some appealing questions of future research are: What are EFL teachers' understanding of formative assessments? What formative assessments do they use to evaluate the four language skills? What follow-up activities do they use after applying formative assessments? Similarly, these research questions can help gather information about the implementation: What facilitates the implementation of formative assessments? What is the relationship(s) among formative assessments, feedback and learning improvement? Finally, to analyze students' points of view, these could be some research questions: What are EFL students' perceptions of formative assessments? What aspects of formative assessment do they appreciate? What causes them to value teachers' feedback more than their peers'?

Challenging Actions

Although many researchers have shown that formative assessments not only help students, but teachers since they guide them to identify strengths and weaknesses and make informed decisions to improve instruction, there is still the need to advocate for their dissemination in EFL classrooms. Therefore, it seems that it is never too late to advocate for their

use to achieve higher English language standards and improve the quality of English instruction in EFL settings. However, as Bronfenbrenner (2001) highlighted, a developmental processes emerge from people and context constant interactions, so it requires to consider characteristics from people, environment, and historical time. Therefore, it can be said that promoting the application of formative assessments in EFL settings may take time and require to start making crucial modifications at the macro, meso, and micro educational levels.

For instance, in non-native English-speaking countries, educational authorities need to start questioning the national assessment policies to find out possible strengths and weaknesses. It will lead to conduct a detailed revision process, make the necessary modifications of national language assessment guidelines, and approve formative assessments as new English assessment practices at the macro level. Then, school administrators need to create a plan to offer teacher training about the new formative assessment practices at the meso level. Finally, when talking about the micro-level, teachers must be willing to incorporate the new learnings acquired from their training program into their daily lessons and make sure students become familiar with them.

All in all, following this process to implement formative assessments in order to create meaningful opportunities for students to improve their English language, and not only domain isolated strategies to pass a test will take time. As it was previously described, it requires to have willingness to improve pre-established policies and traditional teaching and learning assessment practices in EFL classrooms. What is more, it requires having administrators, teachers, and students to work as a team to increase English proficiency levels and have better personal and professional opportunities. As Gallegos (2008) suggested, the English teaching and learning in Ecuador in particular needs to be seen as an educational reform capable of helping citizens to have equitable life opportunities.

References

- Black, P. J., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139-148.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
- Bloom, B. (1984). The search for methods of group instruction as effective as on-to-one tutoring. *Educational Leadership*, 41(8), 4-17.
- Bloom, B. S. (1969). Some theoretical issues relating to educational evaluation. In R. W. Taylor (Ed.), *Educational evaluation: New roles, new means: The 68th yearbook of the National Society for the Study of Evaluation*, Part II (pp. 26-50). Chicago: University of Chicago Press.
- Bronfenbrenner, U. (2001). The bioecological theory of human development. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (pp. 6963- 6970). Oxford, UK: Elsevier.
- Brown, H. D., Abeywickrama, P. (2018). *Language Assessment: Principles and Classroom Practices*. United Kingdom: Pearson Education.
- Burner, T. (2016). Formative assessment of writing in English as a foreign language. *Scandinavian Journal of Educational Research*, 60(6), 626-648.
- Cauley, M. K., & McMillan, H. J. (2010). Formative assessment techniques to support student motivation and achievement. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 83 (1), 1-6.
- Chappuis, S., & Stiggins, R. J. (2002). Classroom assessment for learning. *Educational leadership*, 60(1), 40-44.
- Council of Europe. (2011). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Colley, K. (2008). Performance-based assessment. *Science Teacher*, 75(8), 68-72.
- Foster, D., & Poppers, A. 2009. *Using Formative Assessment to Drive Learning*. Silicon Valley: The Noyce Foundation.

- Gallegos, R. R., Martín, F., Oleas, J., Martínez, D., Minteguiaga, A., Campuzano, A., ... & Ramírez, R. (2008). *Igualmente pobres, desigualmente ricos*. Programa de las Naciones Unidas para el Desarrollo.
- Gordon, E. W., & Rajagopalan, K. (2016). *The Testing and Learning Revolution: The Future of Assessment in Education*. New York, NY: Palgrave Macmillan.
- Gottlieb, M. (2016). *Assessing English language learners: Bridges to educational equity: Connecting academic language proficiency to student achievement*. Corwin Press.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Heritage M. (2010). Formative assessment and next-generation assessment systems: Are we losing an opportunity? *Council of Chief State School Officers*.
- Hirasawa, A. (2013). Self-Assessment and Oral Performance Ability: The Case of Japanese Learners. *Online Submission*.
- Jing Jing, M. (2017). Using Formative Assessment to Facilitate Learner Self-Regulation: A Case Study of Assessment Practices and Student Perceptions in Hong Kong. *Taiwan Journal of TESOL*, 14(1), 87-118.
- Kuzel, M.W., & Shumba, A. 2011. An Investigation into Formative Assessment Practices of Teachers in Selected Schools in Fort Beaufort in South Africa. *J Soc Sci Journal*, 29 (2):159-170.
- Lee, I., & Coniam, D. (2013). Introducing assessment for learning for EFL writing in an assessment of learning examination-driven system in Hong Kong. *Journal of Second Language Writing*, 22(1), 34-50.
- Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: Minute by minute, day by day. *Educational Leadership*, 63(3), 18-24.
- Ministry of Education. (2011). *Acuerdo Ministerial No. 242-11*. Quito, Ecuador: Government Printing Office
- Moss, C. M., & Brookhart, S. M. (2019). *Advancing formative assessment in every classroom: A guide for instructional leaders*. ASCD.

- Naghdi-pour, B. (2017). Incorporating formative assessment in Iranian EFL writing: A case study. *The Curriculum Journal*, 28(2), 283-299.
- Pat-El, R. J., Tillema, H., & Segers, M. (2013). Validation of assessment for learning questionnaires for teachers and students. *British Journal of Educational Psychology*, 83(1), 98-113.
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, 48(1), 4–11.
- Richards, J. C. (2015). *Key Issues in Language Teaching*. Cambridge: Cambridge University Press.
- Ross, S. J. (2005). The impact of assessment method on foreign language proficiency growth. *Applied linguistics*, 26(3), 317-342.
- Ruiz-Primo, M. A. (2011). Informal formative assessment: The role of instructional dialogues in assessing students' learning. *Studies in Educational Evaluation*, 37(1), 15-24.
- Saito, H., & Inoi, S. I. (2017). Junior and senior high school EFL teachers' use of formative assessment: A mixed-methods study. *Language Assessment Quarterly*, 14(3), 213-233.
- Serrano, M. E., Vizcaíno, C. G., Cazco, D., & Kuhlman, N. A. (2015). Transformation: A model for restructuring the preparation of English teachers in Ecuador. *GIST–Education and Learning Research Journal*, (11), 103-117.
- Viktorovna, K. M., & Arkadyevna, S. L. 2015. Formative Assessment as a Component of the Future English Teacher Training. *International Education Studies*, 8 (8): 157-165.
- Widiastuti, I. A. M. S., & Saukah, A. (2017). Formative assessment in efl classroom practices. *Bahasa dan Seni: Jurnal Bahasa, Sastra, Seni, dan Pengajarannya*, 45(1), 50-63.
- William, D. (2010). An integrative summary of the research literature and implications for a new theory of formative assessment. In H. L. Andrade, & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 18–40). New York, NY: Routledge.

Resources for Content-area Teachers Educating ELLs: A Literature Review

Heidi Bartlett

Department of Teaching, Learning, and Teacher Education

University of Nebraska-Lincoln

Abstract

Educating English Language Learners (ELLs) is a rapidly growing need in K-12 schools. While often viewed as a homogeneous group, in reality this population is varied in terms of prior knowledge, access to formal education, age, and native language. Despite these differences, students must be taught first social, and then academic, English in order for them to be successful in their classes and graduate. While in previous decades, ELLs were isolated from the mainstream population, common education practices now integrate them into their grade-level classes. However, while working with this high-need population, mainstream content-area teachers often lack the time, professional knowledge, and/or resources to adequately help. This literature review, focusing primarily on resources for content-area teachers in grades 7-12 from a variety of settings, examines the challenges they face and how some of those challenges can be mitigated. Primarily, the problem must be acknowledged while support is given to teachers to plan and modify their lessons to help older learners who are at basic English proficiency levels. Additionally, students' prior knowledge and experiences must be incorporated into lessons, especially when making connections and in recognizing the funds of knowledge students have. An overarching problem in the field, however, is that studying how to help older learner ELLs is an area that is drastically lacking, and there is definitely room for more focused research on this topic.

Keywords: secondary education, English Language Learners, ELLs, content-area teachers, resources

doi: 10.32873/unl.dc.ne011

Introduction

Students in K-12 English as a Second Language (ESL) programs throughout the United States are a very diverse group. As of Fall 2016, approximately 4.6 million, or 9.6%, of American students are classified as English Language Learners (ELLs) (National Center for Education Statistics, 2019), and they differ in age, native language, literacy skills, amount of formal education, time spent in English-speaking environments, motivation, and family support. Considering the wide range of students that must be educated to function in academic school settings, content-area teachers face an immense challenge in meeting students where they are in order to teach them for academic success. According to the Supreme Court decision *Lau vs. Nichols* in 1974, public schools are required to provide ELLs equal access to education and the accommodations necessary to learn both English and content subjects (Office for Civil Rights: US Department of Education, 2018).

Currently, most states utilize either the ACCESS test through WIDA or the ELPA 21 test through ELPA 21 to check English proficiency annually. Based on these assessments, English Learners (ELs) are either reclassified as English-proficient and exited out of ESL programs under monitoring status, or their scores indicate that they should remain with support for another year. Individual districts and schools are responsible for ensuring that students do not languish in the ESL program and for providing the needed support to exit. Although Annual Yearly Progress no longer needs to be shown to the same degree under the Every Student Succeeds Act (ESSA) as compared to the previous No Child Left Behind legislation, Nebraska has still set a goal that a student entering a school with no English proficiency should exit from the ESL

program within five years. Indeed, Cummins (1981), a noted researcher in the field of second language acquisition, stated, “The finding that it takes at least five years, on the average, for immigrant children who arrive in the host country after the age of six to approach grade norms in L2 CALP [second language Cognitive Academic Language Proficiency] has important educational implications” (p. 148). This conclusion is in line with current Nebraska Department of Education guidelines for ELLs and ESL programs, which is outlined in the Nebraska ELL Program Guide (Nebraska English Learner Programs, 2017).

However, the greater academic load and variety in background knowledge of students entering the upper content courses means that content-area teachers frequently lack the resources to fully support the success of ELs in their courses. Harklau (1994) stated that “... the increased linguistic and academic demands made of older learners make integrated content-area programs, if anything, more crucial for them” (p. 269). Teaching at this upper level requires that students meet a higher standard for exiting ESL programs due to the higher academic and linguistic load. Aside from the use of scaffolding strategies that are generally effective in putting language in context for students, there seems to be little information available about specific curriculum or resources that are both effective for these grades AND that content-area teachers actually utilize. Acknowledging the lack of resources for content area teachers to better engage learning for ELL students is a necessary contribution to the field of education.

This literature review attempts to synthesize research in the field of secondary and post-secondary ESL, with an emphasis on resources available to 7-12 grade content-area teachers. Some studies are content-area specific, while others focus primarily on one of the language domains (such as writing) or culture. Often, researchers included strategies for the use of resources, and those have been included as appropriate.

Reframing Resources

Frequently, when addressing the needs of resources in the classroom, there tends to be a focus on concrete materials such as books and supplemental worksheets. However, this framing of what *resources* entails neglects more intangible factors such as time and space (Wassell, Fernandez Hawrylak, & LaVan, 2010). This narrative research study focused on recent high school graduates who were members of an ESL program at their school and who were enrolled in a bridge program. Researchers examined these intangible resources as provided by field observations and student interviews and concluded that while some of the roadblocks were due to larger educational structures, “teachers who empower ELL students to be productive learners encourage space and time to get to know their students better. This time and space grants students additional access to the curriculum and fosters opportunities for them to speak more English” (Wassell et al., 2010, p. 607). These findings imply that it may be necessary to train content-area teachers on attitudes and beliefs regarding ELLs and the resources that should be considered.

Another type of resource that is not always considered is human capital. To effectively teach English Language Learners, instructors must have not only knowledge of the content, but also understanding of how students acquire language, awareness of areas of potential difficulty, and strategies that can overcome the challenges. However, when trained ESL teachers are not available, the lack of experience and knowledge acts as a roadblock to learning. In one study, “Mainstream teachers were untrained in working with language learners, and ESL and mainstream curricula were not coordinated” (Harklau, 1994, p. 244). Therefore, when considering what type of resources a program has, we need to move beyond the idea of textbooks and technology to include intangible components as well.

Foundational Studies

Many of the foundational studies in TESOL recognize the length of time needed to learn languages in immersion environments. Within the basic idea of a minimum of five years to learn well enough to be successful academically (Cummins, 1981; Hakuta, Butler, & Witt, 2000), there is also an awareness of different types of instruction for older and younger learners.

In Short's (1994) work, she acknowledges the difficulties of older language learners. "Secondary students are at risk, not only because the time available in the K-12 structure runs out sooner for them but also because a large number of the new secondary student immigrating to the U.S. are underprepared for grade-level school work" (p. 582). She recommends that students receive both language and content-area instruction at levels that meet both levels of language and content knowledge (Short, 1994). This can be very difficult, as each secondary student in our schools will arrive with different levels of both, which can make developing standardized materials very time-consuming and challenging. Harklau (1994) supports these ideas, stating, "The content and course objectives of high school subject-area instruction presumed a relatively stable student population with a uniform knowledge base shaped by 8 or 9 years of previous instruction in U.S. elementary and middle schools" (p. 256). This assumption is what causes difficulties for learners in advanced grades as they enter the classroom without the needed knowledge base to be successful.

Indeed, Cummins (1981) relates that school districts may set an arbitrary limit to when services are provided to ELLs that are well below what data states is actually needed. However, since this research about student needs was from 1981, one can hope that more school districts have recognized the importance of providing scaffolded support until students have been deemed proficient enough to succeed academically without support.

In another analysis of different studies (Hakuta et al., 2000), experienced researchers, including Hakuta - a foundational researcher in the field of second language acquisition - investigated previous claims regarding the length of time needed to obtain proficiency. Reviewing the length of time the ESL students attended school in either the US or Canada, the researchers concluded that students needed a minimum of five years to be on par with native speaking students in an academic setting, a conclusion that echoed Cummins' earlier research.

In many states, students have a limited amount of time to be exited from the ESL programs. This study and previous studies from the 1970s to the early 2000s repeatedly show that a minimum of five years is necessary for academic success and fluency. In Nebraska, the current goal is five years to exit an emergent bilingual from an ESL program. However, terms such as *success* and *fluency* need to be operationalized and defined in order to ensure that research is consistent. Additionally, the researchers stated that five years was the minimum length of time necessary and that more time may be needed for some students to obtain fluency. The research also showed that socioeconomic status was a factor in the length of time needed to obtain fluency (Hakuta et al., 2000) which is a factor not always accounted for in research centered on language acquisition.

Content-area Instruction: Social Studies

One direction that studies take is the focus on content-area instruction for language learners. Students frequently face challenges as they are integrated into mainstream classrooms, especially when teachers are unaware of resources and strategies to best assist language learners in acquiring both content and language. However, since ELLs are in mainstream classrooms more frequently than ELL-specific classrooms once they are able to communicate socially, content-

area instruction is a high-need area where teachers need to be aware of the complications when teaching this population.

The content area of social studies seems to be a popular area for research. In one study, Duff (2001) wanted to examine challenges of ESL students in mainstream classrooms and how those challenges were met and focused on the intersection of language, literacy, and culture. In one of two mainstream 10th grade social studies teachers' classes, both native-speaker and non-native-speakers were observed and interviewed for their perspectives about the challenges, and their responses were transcribed and analyzed for themes (Duff, 2001).

One theme that arose when discussing difficulties was regarding resources. In this study, the textbook used in the classes was outdated and written from a white male perspective. Therefore, the teachers supplemented or supplanted the textbook with other readings. Unfortunately, since these readings were collected at the end of class, the students could not refer back to them later or take them home for further study. As the teachers included multiple historical perspectives by inviting guest speakers, classes focused more on oral skills and listening, with few resources such as pictures or diagrams for the ESL students to reference (Duff, 2001). Therefore, even when the social studies teachers strived to make the classes relevant and engaging, many of the students reported that they felt lost during the presentations because of the lack of scaffolding. Although teachers may think they are providing sufficient scaffolding to allow for student comprehension of the topic, students frequently need more than what is provided. Teachers need to think about their course curriculum, resources, and necessary background knowledge in order to help ESL students succeed in their classes.

The resources available to secondary content-area teachers are going to vary by grade and subject. However, among these differences, social studies is a subject area which many ELLs

find interesting due to their life experiences (Short, 1994). In a study that examined the effectiveness of a social studies unit developed for middle school students, Short (1994) enumerated several challenges. One of the primary difficulties was the level of literacy needed to read authentic texts; however, demanding writing requirements, the integrative model of content across grade levels, and the expository textbooks also posed significant challenges to learners (Short, 1994). In her study, these obstacles were addressed through the use of graphic organizers to understand writing and reading organization, the use of supplemental texts to bring the information to a level accessible to students, and cooperative learning tasks to increase communication skills. One advantage that Short (1994) found in her study was that the academic language used in middle school classrooms was frequently on par with other types of academic discourse, with less jargon than might be expected in a subject such as science. From this study, a key conclusion was that language objectives are a requirement when developing lesson plans, and that time, effort, and collaboration are necessary when developing units (Short, 1994). Therefore, when content-area teachers' time or knowledge to develop such effective units and lessons are lacking, our ELL students are often relegated to having to read out of textbooks that are beyond their level and gaining little from the class.

In another study focusing on high school social studies, researchers focused on creating a supportive classroom while reducing cognitive load and explicitly teaching academic skills (Szpara & Ahmad, 2007). This study echoed many of the same difficulties identified by other researchers (Duff, 2001; Short, 1994), but were aided by working with five experienced social studies teachers who were interested in participating in the project. One of the strategies that they focused on was teaching the use of academic resources such as dictionaries (English only) and first language knowledge to rephrase or understand the authentic texts. As translated social

studies textbooks are rare, teachers had to utilize other strategies to reduce the cognitive load of the materials (Szpara & Ahmad, 2007) such as incorporating textbooks from younger grades or guided notetaking. The researchers noted a key difference between ELLs and native-speakers: “Students who speak English as a native language can find and sort key information on their own from either verbal or written sources. In contrast, ELLs’ primary responsibility is to understand new information rather than spend time finding it” (Szpara & Ahmad, 2007, p. 193). This shift in a lesson’s focus requires a change in thinking about the class aims on the part of the content-area teacher, and it must be acknowledged that the ELLs’ primary goal in the classroom is going to be different from a native-speakers’. Both linguistic and content-specific objectives must be developed, addressed, and supported during classes. Again, this requires more professional development for mainstream teachers or collaboration between ESL and content-area teachers.

In another study, Cho and Reich (2008) focused on 33 social studies teachers in six high schools in Virginia, which were target schools for ESL learners. Although ELLs come from a variety of diverse backgrounds and educational histories, many schools use the ELL label as if students were a homogeneous group, which can be problematic. Despite these varied backgrounds, previous research has supported the conclusion that it is beneficial for students to learn both content knowledge and language at the same time. However, some content-area teachers do not agree with this conclusion, as they believe that students would be better placed in isolated language-specific settings where they can receive alternate instruction until they are ready to participate in mainstream classes (Cho & Reich, 2008).

To obtain data from the social studies teachers, Cho and Reich distributed surveys at professional development meetings at the schools and collected them the next day. From the results of the survey, one of the biggest challenges for the social studies teachers was the ELL

students' lack of background knowledge of their area, with 70.6% responding that this one of their top three challenges. In asking about types of classroom supports, many teachers responded that they provided extra time to complete work (Cho & Reich, 2008). However, according to the survey, teachers rarely or never provided alternative assessments or instructional materials to students. Later in the survey, over 75% of the respondents indicated that professional development about ESL instructional strategies would be *very important* or *important* to them. When we look at the results from the non-traditional view of resources (Wassell et al., 2010), we see that the teachers provide some assistance, while lacking in the resources that cost more in time, knowledge, or money (Cho & Reich, 2008).

Regardless, more collaboration with the ESL instructors at the schools, as well as greater communication within departments and between schools, may provide additional support and resources that content-area teachers could develop and provide to students. From a survey like this one, it appears that many teachers claim to be open to receiving more information about how to serve this population better, but may not know where to start in their own classrooms.

Language Domain: Reading

Among all of the different content areas, having literacy skills is a key component that leads to success in secondary grades. When students are not successful in reading, it is less likely that they will find success in content-area classes. In a study concerning the efficacy of reading interventions, Callahan (2010) compared English language development (ELD) curriculum at a high school in California where approximately 37% of the 2000 students at the high school spoke a language other than English at home. With the treatment of the reading intervention program that focused on students in the lowest quartile of reading proficiency, Callahan investigated if the program was effective and if ELLs were being appropriately served

by the new program. As part of the data collection and analysis, long-term ELLs (defined in the study as attending a US school for more than seven years) were also compared to recent immigrants (attending a US school for less than five years) to examine if there were a difference between these two populations and reading program efficacy (Callahan, 2010). Callahan's results showed that long-term ELLs scored worse than recent immigrants in the reading intervention program. More problematically, many of the students who had previously participated in an ELD course were no longer eligible for the reading interventions as they scored too high on the reading proficiency tests, which led to those students not receiving any English support. The differences in test scores, although small, were statistically significant and had the potential to lead to differences in decisions about students (not) being reclassified as ELLs (Callahan, 2010). Therefore, when districts make decisions about academic interventions, they must ensure that the resources being provided are not only useful to teachers, but do not harm the students in their application.

In Callahan's (2010) research, he states the important point, "Learning English does not occur in isolation, rather it is highly context dependent" (p. 3). The idea that English academic instruction is tied to the course content along with instruction about the English language is key to increasing language proficiency. The most effective strategies and resources are going to include both aspect of language learning. Along with considering both aspects, it is vital that the resources available to schools and students be level appropriate and neither too high nor too low.

When focusing on the intersection of reading and ELLs, according to Brown (2007), reading social studies texts specifically places a large burden on students. Social studies texts present a unique challenge in the content area because of its advanced vocabulary, complex syntax, decontextualized information, and assumption of background knowledge. These are

hurdles to overcome for any student not reading at grade level, but challenges are multiplied for ELLs. To combat these, Brown (2007) suggests lexiling texts to make the vocabulary and grammar easier, using guided questions and outlines to keep students on track while reading, and utilizing concept maps to address gaps in knowledge. While not new strategies, these techniques do not address the fact that these scaffolded resources are not immediately available to content-area teachers, and their use is dependent on the teachers' ability to scaffold the written material and time to develop such modifications. Therefore, to be successful, these strategies encompass both the material resources available to teachers as well as the availability of intangible resources.

Language Domain: Culture

Learning about culture is an integral part of learning about language. Without the cultural knowledge of how to use language appropriately in different settings, learning vocabulary and grammar is of limited use. Throughout the world, however, information about the cultural aspect of languages is frequently placed in secondary importance compared to grammar. Specifically, it is asked how international teachers in countries with a national curriculum are implementing cultural knowledge and the degree to which intercultural competence is addressed in the classroom (Lavrenteva & Orland-Barak, 2015).

This study took data from 14 countries with nationalized English curriculums that had information available online for their secondary curricula. The areas that these countries drew from included Asia (2), South America (3), Northern Africa (2), Eastern Europe (4), and Northern Europe (3). The curricula were examined in four dimensions: subject matter, learner, teacher, and milieu. Analysis of the curricula revealed that many countries had similar foci and encouraged that students learn about the attitudes and viewpoints of other cultures while

comparing them to their native cultures (Lavrenteva & Orland-Barak, 2015). Teachers were generally given a lot of leeway in deciding how to implement the curricula and learning objectives.

Unfortunately, this was another empirical study that was not focused on content-area secondary schools and classrooms. However, from this study, we can draw applicable conclusions, most noteworthy that culture is integral to learning a language. The researchers noted that this study lacked information about how the curricular resources and materials addressed culture, so this could be an area of future research. Our content-area teachers will need to consider the cultural aspect to the information they are providing students. Even areas that teachers might view as free from culture, such as math, has cultural components integrated with linguistic components, which need to be addressed and taught to our emergent bilingual students.

Examining culture from another angle, other researchers have examined the relationship between culture and content areas. Lee and Buxton stated, “Because science has traditionally been regarded as culture-free, incorporation of home culture into science instruction has not been adequately conceptualized” (2013, p. 40). This means that science teachers are often unprepared to consider students’ home cultures and what that means for language instruction and the background knowledge that students bring. To improve students’ understanding of science, teachers need to realize that content instruction is not “culture-free”, as everything in education must be situated in context.

Language Domain: Writing

English learners as a population often lag behind in academic writing skills as compared to their oracy skills. In this study, Ramos (2014) worked with a class of 20 public high school

ELLs over a period of eight weeks to improve their writing skills with the *Reading to Learn* curriculum in the northeastern United States. *Reading to Learn* deconstructs texts with teacher assistance and scaffolding to allow students to better manipulate written language. Academic English has very specific writing conventions that are expected in order to be successful; therefore, a persuasive essay was the goal of this unit and this study. The students in this study had scored less in written language skills than in oral language skills, so the goal was to teach specific writing and reading skills have more success later in their high school careers and in postsecondary education (Ramos, 2014).

In the study, pre- and posttests were scored using Functional Language analysis as well as with a more student-friendly tool designed by Ramos titled the Performance Criteria and Assessment Tool (Ramos, 2014). The curriculum was separated into five different stages: Building Field Stage (scaffolding and providing background context), Preparing to Read Stage (providing purpose for a persuasive essay), Detailed Reading Stage (informing about academic writing conventions), Joint Construction Stage (writing a persuasive essay as a class with teacher support), and Individual Construction Stage (student independently writing a persuasive essay). These stages were scaffolded to allow for students to gradually increase their skills and responsibility for their own learning, and posttest results showed that student writing became more complex and organized when compared to their pretest scores (Ramos, 2014).

While this study focused on an ESL classroom without a content-area course as a counterpart, it did provide conclusions about how certain curricula and teaching techniques could be helpful in encouraging academic skills with students. The scaffolded strategies and systematic instruction could provide relevant starting points for addressing learners' needs. However, if content-area teachers do not recognize writing as a set of processes, then they may

not know how to divide the task into its requisite components to encourage student success. Like other areas of language instruction, teaching academic writing requires pedagogical, content, and linguistic skills.

Changing educational standards now focus on ensuring that all students are career and college ready, and this includes students in adult education programs. As our emergent bilinguals are not always able to graduate high school before they are 21 years old, or often drop out before graduating, I expanded the scope of my literature review slightly to examine if post-secondary resources were either more available or more effective in teaching learners. In the U.S. as a whole, the on-time graduation rate in 2016 for ELs was 67% compared to 85% for non-ELs. In Nebraska, the difference in graduation rate was even more drastic, with a rate of 54.5% for ELs and 90.3% for non-ELs – a discrepancy of 35.8% (Department of Education, 2018). These low rates of on-time graduation for language learners clearly indicate that there are deficiencies that are not being addressed in secondary schools.

Writing skills are essential to the education process and life outside of school. This study acknowledged that many adults in community colleges and other post-secondary settings do not finish their degree or program within six years, and this could be partially due to the need for remedial courses at post-secondary levels (Fernandez, Peyton, & Schaetzel, 2017). Adults in such settings come from a wide variety of backgrounds and can include students both with and without a diploma from an American high school. Knowing more about how their high school students will use their knowledge of writing in post-secondary life may help content-area teachers to develop relevant resources and exercises which students will find practical applications for.

Due to these differing needs, Fernandez, Peyton, and Schaetzel (2017) surveyed instructors who taught adult ESL courses and asked them a variety of questions focused on their views of writing importance, types of writing practices, support for writing, and ways to encourage student success. The 43-question survey had 471 participants who completed at least part of the survey. Examination of the results indicated that most students wrote less than one page per week, with lower-proficiency students writing even less. However, respondents indicated that writing was not often tested in determining class placement, and that reading was a larger focus than writing in their programs. Positives to the responses showed that teachers were together with their students regularly, and class sizes were small – both attributes which would be expected to benefit students in a writing class (Fernandez et al., 2017). This is definitely a situation where the instructors may have had the material resources to teach and test writing, but the time and support for this language domain was lacking.

With these results at the postsecondary level, we could expand these conclusions to the contemplation that many ESL students in K-12 settings are not receiving enough writing practice, or concentrated lessons about writing, in their classes due to emphasis on other linguistic skills or content knowledge. Therefore, when students move to content-area classes which require different types of in-depth writing, they may not be sufficiently prepared. ESL teachers and content-area teachers need to examine their curriculum, goals, and expectations to ensure that support is being provided throughout the writing process with appropriate feedback for improvement.

In another study that examined higher-order language use in writing, Mueller and Jacobsen (2015) examined the resources of online dictionaries and corpora to provide students with strategies when they do not know the correct vocabulary to use. Online corpus use has been

a recent area of research in language education as teachers have examined ways to utilize them in classrooms to increase student learning of either vocabulary or correct grammar. These would be especially helpful in the areas of feedback and error correction with written language. This study had two experiments – one examining student perspectives about using an online corpus (n=78) and one investigating if online dictionaries or corpora were more beneficial for collocation and register difficulties (n=39) (Mueller & Jacobsen, 2015).

While the participants in this study were undergraduate females attending a university in Japan taking EFL courses, the ideas concerning the utilization of resources are still applicable to secondary education in the United States. From the first experiment, it was concluded that lower-proficiency learners have difficulty using corpora in an effective manner. From the second experiment, it was determined that students needed more time to learn how to use corpora but were ultimately more successful in gap-fill tests using corpora compared to using an online dictionary (Mueller & Jacobsen, 2015).

A key idea from the results of this study is that instructors need to ensure that appropriate support is provided to ESL students so that they understand how to use the resources and their circumstances for use. Generally, lower-proficiency students have a more difficult time compared to higher-proficiency students when it comes to learning how to utilize language resources. Educators cannot assume that emergent bilinguals are able to use resources effectively. Because of varying skills on the part of students, different types of resources intended to be helpful may actually require more support from the content-area teachers as well as the language-specific teachers in assisting students to learn how to use the resources independently, especially when new.

Conclusion

From the different studies and considerations of the results as it relates to classroom resources, there seems to be a few common themes that would benefit the field of secondary emergent bilinguals. First, content-area teachers frequently lack either the knowledge or the time (or both) to modify existing resources such as textbooks to better suit the needs of our language learners. Second, secondary students (grades 7-12) who enter as newcomers are at a disadvantage not only due to language constraints, but also due to assumptions about common life experiences and the pressure of earning credits toward graduation quickly. When content-area teachers do not think about or plan for the assumed cultural knowledge needed to be successful in a task, it increases the already heavy linguistic burden that students carry.

Third, systematic support is needed both for emergent bilinguals AND content-area teachers. For students, support could mean pre-teaching vocabulary, learning how to write extended discourse in English, or safe spaces to practice classroom speaking. For mainstream teachers, support could be modeling effective teaching with strategies such as utilizing graphic organizers, conversations with the ELL team to discuss specific student needs and experiences, or learning how to rewrite texts to be level-appropriate. Regardless of what the support looks like in schools, support is both federally mandated for students AND leads to better learning outcomes when people are able to access the help they need.

However, in analyzing the research concerning where 7-12 grade ELLs need the most resources and support, it has become obvious that this entire field is lacking in empirical research. From what studies we do have, it is clear that support and resources are underprovided when teaching our older language learners. This could be due to several reasons. For one, students who are classified as ELL in kindergarten are often exited out of an ELL program by the

time they reach high school. As students enter ELL programs at an older age, the gap between their language skills and what their grade-level peers are learning in their content areas is wide. Failing adequate resources and education on the topic, content-area teachers may neglect to support these students in ways that are beneficial for decreasing that gap and for future learning.

Another point is that not only does there need to be more research in general, there also needs to be more research in content areas other than social studies. I'm not sure why this content area seems to be more prevalent than science or English Language Arts, for example. However, it could be due to multiple factors, not least of which that many people consider social studies to be less important than math or reading (as an "untested" subject), and could therefore be easier to receive permission from school districts for research in this content area.

Finally, where there is research, it tends to focus on specific content areas such as social studies, with others, such as math, lacking in research. Due to the paucity of research, any type of empirical study that concentrates on secondary ELLs would benefit the field of teaching English to speakers of other languages. Areas for further research could take a variety of directions: successful strategies utilized in classes with ELLs by content-area teachers and by ELL teachers, longitudinal studies which follow the success (or not) of ELLs in order to find factors which lead to timely high school graduation, textbooks or supplemental resources that are scaffolded and lexiled for learning, or development of resources that are useful to an emergent bilingual population that has rich life experiences which don't match their English language ability. Any study of this type could help provide starting points in expanding this field to encompass the great variety of students and situations that we serve. Additionally, Wassell et al. noted that much of the current research is focused on deficit views of teaching language learners;

teachers need to be aware of this approach towards research in order to acknowledge a needed shift in research (2010).

In K-12 education, we want our students to be prepared for their later life goals. Language skills are an important component of this goal. As teachers, we need to recognize the challenges and strengths that language learners bring to our classrooms while working with students to overcome weaknesses. While working with emergent bilinguals may be unfamiliar to some content-area teachers, it is important that teachers educate themselves about their needs and work to integrate teaching strategies that can compensate for developing English skills. As educators, we must ensure that student life goals are met and that our students are supported along the way to success.

References

- Brown, C. L. (2007). Strategies for making social studies texts more comprehensible for English-language learners. *The Social Studies*, (October), 185–188.
- Callahan, R. M. (2010). The intersection of accountability and language: Can reading intervention replace English language development? *Bilingual Research Journal*, 30(1), 1–21. <https://doi.org/10.1080/15235882.2006.10162863>
- Cho, S., & Reich, G. A. (2008). New immigrants, new challenges: High school social studies teachers and English language learner instruction. *The Social Studies*, (November/December), 235–242.
- Cummins, J. (1981). Age on arrival and immigrant second language learning in Canada: A reassessment. *Applied Linguistics*, 11(2), 132–149. <https://doi.org/10.1093/applin/2.2.132>
- Department of Education. (February 2018). *Academic performance and outcomes for English learners*. Retrieved from <https://www2.ed.gov/datastory/el-outcomes/index.html#two>
- Duff, P. A. (2001). Language, literacy, content, and (pop) culture: Challenges for ESL students in mainstream courses. *The Canadian Modern Languages Review/La Revue Canadienne Des Langues Vivantes*, 58(1), 103–132.
- Fernandez, R., Peyton, J. K., & Schaetzel, K. (2017). A survey of writing instruction in adult ESL programs: Are teaching practices meeting adult learner needs? *Journal of Research and Practice for Adult Literacy, Secondary, and Basic Education*, 6(2), 5–20.
- Hakuta, K., Butler, Y. G., & Witt, D. (2000). *How long does it take English learners to attain proficiency?*
- Harklau, L. (1994). ESL versus mainstream classes: Contrasting L2 learning environments. *TESOL Quarterly*, 28(2), 241–272.
- Lavrenteva, E., & Orland-Barak, L. (2015). The treatment of culture in the foreign language curriculum: An analysis of national curriculum documents. *Journal of Curriculum Studies*, 47(5), 653–684. <https://doi.org/http://dx.doi.org/10.1080/00220272.2015.1056233>
- Lee, O., & Buxton, C. A. (2013). Integrating science And English proficiency for English language learners. *Theory Into Practice*, 52, 36–42.

<https://doi.org/10.1080/07351690.2013.743772>

- Mueller, C. M., & Jacobsen, N. D. (2015). A comparison of the effectiveness of EFL students' use of dictionaries and an online corpus for the enhancement of revision skills. *ReCALL*, 28(1), 3–21. <https://doi.org/10.1017/S0958344015000142>
- National Center for Education Statistics. (May 2019). *English language learners in public schools*. Retrieved from https://nces.ed.gov/programs/coe/indicator_cgf.asp#info
- Nebraska English Learner Programs. (2017). *Nebraska ELL program guide: Guide for teachers*. Nebraska Department of Education. <https://cdn.education.ne.gov/wp-content/uploads/2017/10/NDE-Teacher-Program-Guide-2017-updates.pdf>
- Office for Civil Rights: US Department of Education. (25 September 2018). *Developing programs for English language learners: Lau v. Nichols*. Retrieved from <https://www2.ed.gov/about/offices/list/ocr/ell/lau.html>
- Ramos, K. (2014). Teaching adolescent ELs to write academic-style persuasive essays. *Journal of Adolescent & Adult Literacy*, 57(8), 655–665. <https://doi.org/10.1002/jaal.303>
- Short, D. J. (1994). Expanding middle school horizons: Integrating language, culture, and social studies. *TESOL Quarterly*, 28(3), 581–608.
- Szpara, M. Y., & Ahmad, I. (2007). Supporting English-language learners in social studies class: Results from a study of high school teachers. *The Social Studies*, (October), 187–195.
- Wassell, B. A., Fernandez Hawrylak, M., & LaVan, S.-K. (2010). Examining the structures that impact English language learners' agency in urban high schools: Resources and roadblocks in the classroom. *Education and Urban Society*, 42(5), 599–619. <https://doi.org/10.1177/0013124510375598>